Inter-generational Congestion and Spatial Labour Market Outcomes*

Gabriele Guaitoli[1], Roberto Pancrazi[2], and Damiano Raimondo[2]

[1]Department of Applied Economics, Universitat Autonoma de Barcelona

[2]Department of Economics, University of Warwick

March 23, 2026

Abstract

This paper studies how age-varying housing costs arising from borrowing constraints shape the spatial sorting of workers and aggregate welfare. Using U.K. data, we document that younger homeowners pay 10–15 percent more in housing costs than older workers, everything else equal, primarily due to higher loan-to-value ratios, which explain 35–40 percent of this gap. This cost differential is larger in high-productivity regions offering steep early-career wage growth. Consistent with these frictions, high-wage labour markets have aged disproportionately over the past two decades. We develop a spatial overlapping generations model where intergenerational congestion arises as higher housing costs exclude young workers from productive areas offering the highest returns to human capital accumulation. The market equilibrium is inefficient, generating both a static sorting wedge and a dynamic human capital wedge. Extending the model to a quantitative spatial framework with 52 UK Travel-to-Work Areas, heterogeneous agents, and endogenous homeownership choices, we find substantial welfare losses from intergenerational congestion. We evaluate a revenue-neutral place-age-based council tax reform that reduces housing costs for young workers (aged 21–34) in productive locations, offset by a modest increase for older workers (aged 40+). The gap-equalising policy raises output by nearly 0.1% and welfare by 0.3% in consumption-equivalent terms; a slightly larger subsidy doubles these gains. These improvements arise from better age sorting within productive locations.

JEL Classification: R23, R31, J24, E24, G21, G51

*The authors declare no sources of external research support and no financial relationships or other potential conflicts of interest.

1 Introduction

Over the past two decades, a striking pattern has emerged in England and Wales's landscape: the most productive locations, those offering both high wages and a fast accumulation of human capital (la Roca and Puga, 2017; Lhuillier, 2024; Acabbi et al., 2024), have aged substantially faster than the rest of the country. Between 2002 and 2019, commuting zones in the top wage decile saw the mean age of working-age residents increase by 1.2 years, nearly double the national average of 0.7. This differential is driven by a steep decline in the share of young workers (aged 25-34) in high-wage locations (-2.0 percentage points, from the initial 25 percent share in 2002), compared to a mere 0.3 p.p. decline nationally.[1]

This demographic shift raises a fundamental economic question: is the age composition of productive locations efficient? When young and old workers compete for limited space and housing in productive locations, creating what we call *intergenerational congestion*, human capital could be misallocated if the young are priced out from good jobs and learning opportunities.

We show that housing market frictions are a key driver of spatial age misallocation. Younger workers pay 13-19% more in housing costs than older workers to live in the same location, in similar properties. Loan-to-value differences, reflecting young workers' limited wealth and consequent higher debt leverage and costs (Stein, 1995), explain 35-40% of this gap. Critically, the gap is larger in high-productivity, high-learning locations. This friction makes the *intergenerational congestion* arising from competition for urban space inefficient: older workers, facing lower housing costs, crowd out younger workers from productive locations, generating both a static misallocation (due to insufficient same-period sorting over skills) and dynamic losses (due to insufficient human capital accumulation) that persist across generations.

We show this result in three steps. First, we establish stlyzed empirical facts using UK microdata on housing costs, wages, and demographic composition across local labor markets. Second, we develop a tractable spatial overlapping generations model to derive the theoretical conditions under which age-varying housing costs generate inefficient spatial sorting. Third, we extend the framework to a quantitative spatial model calibrated to 52 UK Travel-to-Work Areas (TTWAs) to measure the magnitude of welfare losses from intergenerational congestion and evaluate policy interventions.

We provide new empirical evidence for this housing cost channel using UK data, connecting and extending existing findings on mortgage market frictions. First, building on the well-documented relationship between loan-to-value ratios and mortgage costs (Stein, 1995; Bester, 1985; Adams et al., 2009), we use detailed micro-data from the English Housing Survey to show that younger workers pay 13-19% more in housing costs than older workers for compa-

[1] The top-decile wage locations include 32 percent of the overall working age population.

rable housing in the same locations, even after accounting for the opportunity cost of the initial downpayment. Consistent with the asymmetric information channel documented by Admati and Hellwig (2014) and Benetton et al. (2018), we find that loan-to-value differences, reflecting younger workers' limited wealth, explain 35-40% of this age gap. We document that this differential is most pronounced in productive, expensive regions. Second, we establish that productive TTWAs offer not only higher wages but also steeper early-career wage growth, confirming and extending the urban learning premium (la Roca and Puga, 2017; Glaeser and Maré, 2001; Lhuillier, 2024; Acabbi et al., 2024). This spatial heterogeneity in returns to experience implies that limited access to these labor markets is particularly costly for the young, who risk missing a crucial phase of human capital accumulation. Third, we document a novel fact: high-wage, high-wage-growth labor markets in England have aged disproportionately faster over the past two decades. A 10-percent difference in wages offered across TTWAs is associated with roughly a 0.6 percentage point *faster* increase in the relative shares of older and younger workers between 2002 and 2019.

To assess whether the observed sorting patterns are efficient, we develop a tractable two-location spatial overlapping generations model. Workers live for two periods, choosing their location each period based on wages, amenities, and housing costs. Locations differ in productivity and returns to experience. The key friction is that young and old workers face different housing costs in the same location, which we initially treat as exogenous. Importantly, we assume that there are no other borrowing constraints and frictions to spatial mobility. We show that the housing market friction alone generates what we call *intergenerational congestion*: when generations compete for scarce space - albeit for different, yet correlated, incentives- but face different prices, the market equilibrium is inefficient. Old workers, valuing only current wages, crowd out young workers who value both current wages and future skill accumulation. The inefficiency operates through two channels. The first one is a *static sorting wedge*: since old workers face lower housing costs, they outbid young workers for city locations even when the young would derive higher utility from living there. This wedge measures the potential current-utility gain from efficient reallocation. The second one is a *dynamic human capital wedge*, which captures the foregone option value of experience when young workers are priced out of skill-enhancing cities. We characterize the planner's allocation and show it features more young workers in productive locations than the market equilibrium. The welfare loss from intergenerational congestion increases with the housing cost differential between young and old and with the spatial concentration of both productivity and learning opportunities.

We extend the theoretical framework to a quantitative spatial model to measure welfare losses and evaluate policies. The quantitative model features 52 UK Travel-to-Work Areas, a realistic life-cycle structure with multiple working periods and retirement, and ex-ante heterogeneity in workers' skills and housing wealth. We endogenize homeownership decisions and optimal

deposit choices, which microfound the housing cost differential between young and old workers, and incorporate location-specific skill accumulation with peer learning effects and realistic moving costs. We estimate the model via Simulated Method of Moments using microdata on wages, housing costs, migration, and demographics across UK labor markets, solving for the spatial equilibrium on a fine grid of locations and worker types. The estimated model successfully replicates key spatial patterns in the data.

We then evaluate a revenue-neutral place-age-based council tax reform that reduces housing costs for young workers (aged 21–34) in productive locations, offset by a modest increase for older workers (aged 40+). We find that equalising the housing-cost age gap across productive and unproductive locations raises output by nearly 0.1% and delivers a welfare gain of 0.3% in consumption-equivalent terms; a larger subsidy of 2% of the house value doubles these gains. These improvements arise not from moving more workers into productive locations — whose overall population share slightly falls — but from improving the age composition within them: the share of young workers in productive locations rises, stimulating skill accumulation and raising aggregate output. Finally, we show that these gains are not driven by incentivising the relocation of retirees: excluding pensioners from the tax increase preserves between 75% and 90% of the output gains and all welfare gains, confirming that the key distortion is intergenerational competition for space among workers with different learning rates, rather than competition between productive and non-productive individuals.

The remainder of the paper is organized as follows. Section 2 documents the empirical facts on housing costs, wages, and aging across UK labor markets. Section 3 develops the theoretical model and derives conditions for inefficiency. Section 4 presents the quantitative model and calibration. Section 5 and Section 6 reports the main estimation results and policy evaluation. Section 7 concludes.

Related Literature

Our paper bridges three literatures: household finance and housing markets, spatial economics and agglomeration, and overlapping generations models.

Housing finance and life-cycle considerations. Research on financial frictions has documented their effects on household decisions and aggregate outcomes (Kaplan and Violante, 2014; Kaplan et al., 2014; Huggett and Ventura, 1999). In housing markets, the relationship between loan-to-value ratios and mortgage costs is well established (Stein, 1995; Bester, 1985; Adams et al., 2009; Liu, 2022), with LTV premia rationalized by asymmetric information (Admati and Hellwig, 2014; Benetton et al., 2018). Recent work emphasizes how these frictions shape consumption, homeownership timing, and wealth accumulation over the life cycle (Attanasio et al.,

2012; Ortalo-Magné and Rady, 2006; Kaplan et al., 2020). We contribute by introducing a novel channel through which these frictions misallocate labor across space: age-varying housing costs distort the spatial allocation of young workers to high-wage growth, high-wage locations, reducing aggregate productivity not only via a static misallocation, but also a dynamic loss of human capital creation.

Spatial economics and urban productivity. Cities generate productivity advantages through better matching (Moretti, 2011), knowledge spillovers (Glaeser and Maré, 2001), and faster human capital accumulation (la Roca and Puga, 2017). Recent work quantifies substantial aggregate welfare losses from housing supply constraints limiting productive cities' *size* (Hsieh and Moretti, 2019; Ganong and Shoag, 2017). We complement this by identifying an orthogonal source of misallocation: the age *sorting* across cities is inefficient, as too many young workers are priced out of high-learning locations, even when size is fixed. Related work examines how housing costs interact with sorting (Moretti, 2012; Diamond, 2016; Ganong and Shoag, 2017), but we highlight the intergenerational dimension and the aggregate implications of individuals of different ages and skills being differently exposed to housing market frictions. In a similar fashion, Lhuillier (2024) studies how learning complementarities among high-skilled individuals imply that cities may not be skilled, rather than large, enough.

Overlapping generations models. We build on classic OLG frameworks (Samuelson, 1958; Diamond, 1965b), introducing spatial heterogeneity to show age-varying housing costs create a novel inefficiency channel. When generations compete for urban space but face different prices and location incentives, the market features too many old and too few young workers in productive locations. This intergenerational congestion generates both static misallocation and dynamic human capital losses, arising even without standard OLG frictions like borrowing constraints or missing markets.

Our quantitative framework builds on advances in quantitative spatial economics (Greaney et al., 2025) and models with skill accumulation (Lhuillier, 2024). We introduce age-heterogeneous housing costs arising from financial frictions, central to our mechanism. Our policy analysis contributes to debates on place-based policies (Fajgelbaum and Gaubert, 2020), showing that age-place-based interventions reduce welfare losses while mitigating spatial and intergenerational inequality.

2 Housing costs, Aging, and Productivity in the UK

Mortgage market frictions arising from limited liability and default risk fundamentally shape the financial conditions available to prospective homebuyers (Piazzesi and Schneider, 2016; Benetton et al., 2018; Bester, 1985; Stein, 1995). Because lenders incorporate default probabilities

into mortgage contracts, borrowers' wealth positions critically determine their access to credit, leverage opportunities, and loan terms. Younger households, who typically possess limited wealth, consequently face tighter borrowing constraints and less favorable mortgage conditions than their older, wealthier counterparts (Haurin et al., 1996; Ortalo-Magné and Rady, 2006; Liu, 2022). While prior research has emphasized how these frictions affect aggregate consumption and homeownership patterns (Kaplan and Violante, 2014; Kaplan et al., 2020), their implications for the spatial allocation of labor and the resulting aggregate productivity consequences remain unexplored. We document three novel empirical facts that reveal this spatial dimension. First, younger workers face systematically higher housing costs than older workers in the same locations, with differences in loan-to-value ratios and associated interest rate premia accounting for a substantial portion of this gap (Fact 1). Second, housing cost differentials between age groups are most pronounced in high-productivity regions—precisely the locations where younger workers stand to gain the largest returns to experience through accelerated human capital accumulation (De la Roca and Puga, 2017; Glaeser and Maré, 2001; Acabbi et al., 2024) (Fact 2). Third, these high-productivity areas have experienced disproportionately rapid workforce aging over the past two decades (Lee and Sissons, 2016; Champion, 2016) (Fact 3).

2.1 Fact 1: Age-specific housing costs

In this subsection we provide evidence that housing costs for mortgagers are higher for the young and lower for the old.[2] A sizeable share of this age-gap in housing costs is captured by differences in Loan-To-Value ratios (LTVs) across age groups. To assess this point we utilize data from the English Housing Survey (EHS), a long-running, representative survey that offers detailed insights into housing stock, prices and household characteristics in England.[3]

The EHS collects information on housing costs (rents and mortgage payments), conditions, tenure, affordability, energy efficiency, and demographic characteristics. While the whole dataset covers the period 1993-2022, we restrict our analysis to the period 2009-2013, during which the survey reports information on price paid, deposits, and current value of the properties. This information allows us to build a comprehensive housing cost variable, which includes direct payments, maintenance, and the opportunity cost of the initial deposit. Hence, we are able to compare the shadow costs faced by households who bought equally-valued houses with different leverage. Our final sample contains 67,268 observations across renters, mortgagers, and owners outright. Appendix A.1 presents descriptive statistics of the dataset.

We first display the distribution of the Loan-To-Value ratios (LTVs) across different age groups

[2] In Appendix A.2 we show that all the results hold when considering mortgage-less house owners, as well as renters.

[3] The EHS has been conducted continuously since 2008, following its merger with the Survey of English Housing (SEH), which was launched in 1993.

in Table 1. LTVs are calculated as the downpayment paid for the house divided by the price paid, capturing the LTV at mortgage origination. A LTV lower than 60% is considered the safest category of loans, while LTVs larger than 90% are considered particularly risky. When the house was purchased without a mortgage, we denote it as a "Cash" purchase. Around half of homeowners aged 45 to 64 years old purchased their house either cash or with LTVs lower than 60%. In contrast, only 11% of young households aged 25-34 falls into this category. Moreover, over one quarter of young households borrowed with 90% or larger LTVs, compared to just 13% of older ones. Nearly two-thirds of young homebuyers had LTVs higher than 80%.

Table 1: Share of homeonwers in each LTV category at mortgage origination, by current age group

	LTV Ratio					
Age Group	Cash	$\leq 60\%$	60-70%	70-80%	80-90%	$\geq 90\%$
25-34	0.029	0.088	0.077	0.178	0.364	0.264
35-44	0.040	0.193	0.113	0.163	0.270	0.221
45-64	0.144	0.299	0.110	0.127	0.187	0.133
65+	0.474	0.211	0.046	0.068	0.123	0.079

Notes: The table shows the share of homeowners in each LTV category at the time of mortgage origination, by age group, based on survey data from 2009 to 2013. Households are weighted by the number of household heads. The shares are calculated based on the weights provided in the dataset. The LTV categories are defined based on the ratio of the deposit to the total value of the property financed by the mortgage, at the moment of purchase. We do not report further LTV subdivisions below 60%, as this is considered the safest loan category by most banks.

Higher loan-to-value (LTV) ratios are consistently associated with higher mortgage interest rates, reflecting the increased credit risk that lenders face when borrowers contribute less equity.[4] For example, in May 2025 Bank of England reported that the average interest rate quoted for 2-years fixed 95% LTV loans had a 1.0 percentage point spread against 60% LTV loans (5.09% vs 4.09%). This gap corresponds to a 10% larger monthly payment for a 25-years loan of identical size. In December 2014, the average interest rate for 95% LTV loans was almost three times the one for 60% LTV loans (5.27% vs 1.90%), implying a 42% larger monthly payment for a 25-years loan.

Do the higher LTV for the young correspond to higher housing costs? We present evidence of a novel fact that has been largely overlooked in the literature: holding other factors constant, younger households face higher housing costs than older ones. Our variable of interest is housing cost, measured as a share of the initial price paid. To account for the opportunity cost of the downpayment and the depreciation of the asset, we compute housing cost as the sum of: 1) the mortgage costs reported by households, C_{it}; 2) $r \times (1 - LTV_i) \times (\text{House Value})_{it}$, representing

[4]See (Liu, 2022; Qi et al., 2009; Dell'Ariccia et al., 2012). These studies provide robust theoretical and empirical support for the practice of charging higher mortgage interest rates as LTV increases.

the shadow cost of not having invested the deposit in a low-risk alternative investment, which pays dividends per year at the net interest r and which we assume would have appreciated (on top of the dividend) at the same pace as the house itself;[5] 3) a maintenance cost equal to the average depreciation of housing assets, as a share of land plus building value δ of the current house value $(\text{House Value})_t$, from UK national accounts. We then normalise all these costs as yearly shares of the initial housing cost, so that our cost index is:

$$H_{it} = \frac{C_{it} + [\delta + r \times (1 - LTV_i)] \times (\text{House Value})_{it}}{(\text{Initial Value})_i}.$$

We assume a risk-free real interest rate equal to 2% , $r = 0.02$, and a housing depreciation of 0.895%, $\sigma = 0.00895$, calculated from UK national accounts.[6] Finally, In order to exclude clear cases of remortgaging, which are not necessarily connected to house purchasing, we exclude all observations of mortgagers for which the initial mortgage taken to buy the house should have ended before the survey date, or that explicitly state their current mortgage is not connected to the purchase.

To corroborate the evidence of age-specific housing costs, we estimate the following regression model:

$$\log(H_{i,j,t}) = \gamma^g Age^g_{i,j,t} + \beta_1 Term_{i,j} + \beta_2 \log(Inc_{i,j,t}) + \beta_4 LTV_{i,j} + \alpha_j + \theta_t + \varepsilon_{i,j,t},$$

where $H_{i,j,t}$ denotes the housing cost of household i residing in location j in year t, as explained above; $Inc_{i,j,t}$ is the joint household income; $Term_{i,j}$ indicates the length of the mortgage, capturing that identical, longer mortgages have lower yearly costs; $LTV_{i,j}$ are variables related total amount of deposit placed; α_j are location fixed effects (nine regions in England and Wales); θ_t are year fixed effects. The coefficients of interest, γ^g, captures the average normalized housing cost for each age group g. The reference group is young workers aged 25–34, so that $Age^g \in \{35\text{–}44, 45\text{–}64\}$. Table 2 reports the regression results. Column (1) presents the baseline specification with region, year, and year-of-purchase fixed effects. The latter allows to neutralize potential yearly differences in average mortgage costs (e.g. due to interest rate changes), allowing to compare purchases made in different years. The coefficients on the age-group indicators are negative and decreasing in age, showing that older workers face sys-

[5] Notice that, by investing the deposit in a similar asset, it would have appreciated at the same pace as the house value. Hence, the dividend paid at time t of a (1-LTV)% deposit paid for a house of price P_{t-n} at time $t-n$ must be the one of a (1-LTV)% deposit paid for a house of price $P_t = (\text{House Value})_{it}$ at time t. We perform robustness checks regarding the dividend size in Appendix A.2.2. The rental yield of an average UK house is estimated to be 3.5%, net of taxes and costs but not accounting periods of potential lack of income (e.g., extraordinary repairs, gaps between rental contracts, non-payment and disputes) and the risk premium and time costs (e.g., managing the property, being in touch with the agent, ...) associated with having invested in a single property, rather than a financial portfolio.

[6] See Appendix C.2 for details.

tematically lower housing costs relative to young workers, the omitted reference group. On average, older workers pay about 14% less. Specification (2) controls for mortgage duration. The cost of mortgages, in fact, depends on their duration as longer terms are associated with lower monthly payments. Controlling for the duration increases the gap to 15%. In column (3) we control for income, which does not affect the results.[7] Column (4) interacts the age-group dummy with a London–South East dummy, the most productive and expensive areas in England. The results reinforce the baseline findings and reveal that the housing cost penalty for young workers is even more pronounced in these locations, by an additional 7%. Column (5) introduces the percentage of the house value deposited as downpayment. As expected, higher deposits reduce the cost. Moreover, since the young pay smaller average downpayments, the young-old cost gap is reduced to 10 p.p., meaning that downpayment differences explain 35 percent of the cost gap. In column (6) we add the age groups-locations interactions with the deposit control. Also in this case, the deposit percentage explains 40 percent of the age cost gap. Finally, in column (7) we replace the continuous deposit percentage variable with discrete categories that indicate the mortgage's LTV ratio. As expected, mortgage costs are monotonically increasing in the LTV ratio categories, which explain 35 percent of the age-cost gap.

In Appendix A.2 we provide further robustness checks around the definition of house costs, and show similar facts for renters.

[7] This is in line with expectations. First, we normalise costs by the value of the property. Second, controlling for mortgage duration further accounts for the income of a household, with lower-income households seeking longer mortgages to lower the monthly payment.

Table 2: Age-specific mortgage costs

	ln(opportunity cost)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age 35-44	-0.083***	-0.088***	-0.089***	-0.075***	-0.071***	-0.055***	-0.056***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Age 45-64	-0.136***	-0.148***	-0.149***	-0.127***	-0.099***	-0.077***	-0.083***
	(0.013)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)
ln(Mortgage years)		-0.035*	-0.035*	-0.034*	-0.111***	-0.110***	-0.082***
		(0.020)	(0.020)	(0.020)	(0.022)	(0.022)	(0.021)
ln(income/head)			0.014				
			(0.011)				
Age 35-44 × (London+SE)				-0.045*		-0.053*	-0.052*
				(0.027)		(0.027)	(0.027)
Age 45-64 × (London+SE)				-0.066**		-0.070**	-0.067**
				(0.028)		(0.028)	(0.028)
Deposit share					-0.369***	-0.369***	
					(0.027)	(0.027)	
LTV <60%							-0.170***
							(0.017)
LTV 60-70%							-0.096***
							(0.018)
LTV 70-80%							-0.069***
							(0.016)
LTV 80-90%							-0.027*
							(0.015)
Observations	8,107	8,107	8,107	8,107	8,078	8,078	8,078
R^2	0.410	0.410	0.410	0.411	0.429	0.429	0.424
Within Adjusted R^2	0.015	0.015	0.015	0.016	0.046	0.047	0.039
Region fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓
Year Purchased fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: $^{***} \leq 0.01$, $^{**} \leq 0.05$, $^{*} \leq 0.10$. . The dependent variable is the imputed opportunity cost of the mortgage, which includes yearly costs plus the shadow cost of the capital used as deposit, which is assumed to be invested in an identical property (and, hence, is revaluated at the same rate of the property's value) with a yearly real yield, on top of re-evaluation reinvested in the property itself, of 2%. The opportunity cost also includes the cost of maintenace, estimated to be 0.895% of the current house value, per year. The cost is then normalised by the initial value of the property. Heteroskedasticity-consistent (HC) standard errors are reported in parentheses.

2.2 Fact 2: Age-specific location return

In the previous section, we showed that housing costs are different between young and old workers. The difference is particularly large in London and the surrounding South-East region, which are the most expansive and productive in the UK. In this subsection, we highlight that those areas are characterized by labor markets that provide, at the same time, higher wages for all workers - due to their higher productivity - as well as larger return to experience for young workers.

To investigate this, we draw on microdata from the Annual Survey of Hours and Earnings (ASHE), a matched employer-employee dataset for Great Britain. ASHE includes detailed information on earnings, occupation, demographic characteristics of workers and firms, and

precise geo-coding of both workplace and residence locations, based on a 1% panel sample of all employees.[8] We restrict our analysis to the period from 2004 to 2019[9], and exclude workers that appear in the panel only in once, bringing the total number of observations to $2,251,607$. For each subsequent analysis, we use the appropriate calibration weight to match official aggregate demographic and labour market statistics. Among the information on earnings we use hourly earnings, derived as the ratio between gross pay and total paid hours (including overtime). Finally, following Lhuillier (2024), we normalize wages to have the same mean in each period. This normalisation allows us to abstract from aggregate wage growth when studying the evolution of wages over the life cycle.

First, we compute the age-specific average wage in each Travel to work areas (TTWA).[10] To do so, we define as *Stayers* those individuals who have lived and worked in the same area for at least three consecutive years.[11] Formally, let g denote an age group, i a location, and t a year. The *average wage of stayer workers* in age group g, location j, and year t, denoted by $\text{AvgWage}_{j,t}^{g,\text{stayer}}$, is calculated as the average hourly earnings in year t among individuals who (i) belong to age group g in year t, (ii) have been continuously observed in location j over the previous $k = \min(3, t - t_0 + 1)$ years, where t_0 is the initial year in the sample, i.e. 2004; and (iii) have remained within age group g over the same period.

Let $\mathscr{I}_{j,t}^{g}$ denote the set of individuals i who satisfy the above conditions—i.e., those observed in location i and age group g for all years in the interval $[t-k+1, t]$. Then the average wage is given by:

$$\text{AvgWage}_{j}^{g,\text{stayer}} = \frac{1}{16} \sum_{t=2004}^{2019} \frac{\sum_{i \in \mathscr{I}_{j,t}^{g}} \omega_{i,t} \cdot \text{he}_{i,t}}{\sum_{i \in \mathscr{I}_{j,t}^{g}} \omega_{i,t}},$$

where $\omega_{i,t}$ denotes the sampling weight for individual i in year t and $he_{i,t}$ the normalised hourly earnings.

Figure 1 displays the average log wage over the period 2004–2019 for both young and older workers. While, as expected, average wages are higher for older individuals—consistent with

[8] The sample is selected by including all individuals whose National Insurance Number ends with two specific digits. Information is reported for all individuals selected this way, who were employed in April of the year. The data include administrative data from HMRC, the tax agency, and complementary data, such as precise occupational codes, filled in via a survey sent to the employer. Employers have a legal obligation to answer to such request for information.

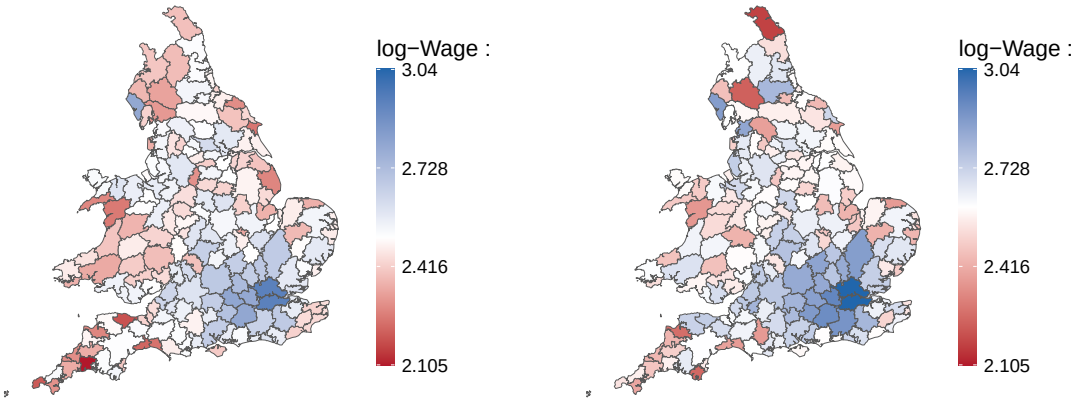
[9] Before 2004, ASHE lacks appropriate sampling weights, making earlier years unsuitable for panel analyses.

[10] Travel to Work Areas are defined by the Office for National Statistics (ONS) as self-contained labor market areas where at least 75% of the resident population works locally and at least 75% of the local workforce resides in the same area. These units are constructed using commuting flow data. There are a total of 173 TTWAs for England and Wales.

[11] This rapresents around 56% of our total sample of workers (310,224). If we also include workers with employment gaps within the horizon, the percentage increases to approximately 70%.

standard life-cycle earnings profiles—the spatial distribution of wages is remarkably similar across age groups. In particular, wages are highest in and around London and in the South-East of England, which are also the most productive regions. Unsurprisingly, the correlation between average wages for young and old worker is quite high, as illustrated by Figure 3a.

Figure 1: Average Wage for stayers by age



(a) Young (b) Old

Notes: The figure shows the average log real hourly earnings for individuals who did not leave the workforce or changed location over a period of three years ("stayers"). Panel (a) reports the figure for individuals aged 25-34, and panel (b) for individuals aged 50-64 years old. Locations with wages above the age-specific median are depicted in increasingly darker shades of blue, and those below in increasingly darker shades of red. Those around the age-specific median are in white.

Second, we compute a location-specific measure of wage growth. For each individual stayer we compute the annualized real hourly earnings growth over a 3 year horizon:

$$WageGrowth_{i.t} = \left(\frac{he_{i,t+3}}{he_{i,t}} \right)^{1/3} - 1.$$

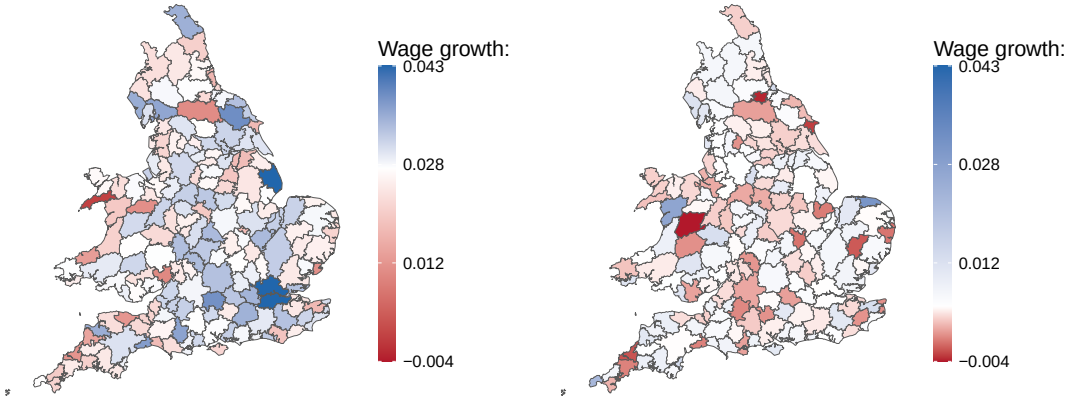
Then, for each TTWA (j) we define the average wage growth for each age-group g:[12]

$$(AvgWageGrowth)_j^{g,stayer} = \frac{1}{14} \sum_{t=2006}^{2019} \frac{\sum_{i \in \mathscr{J}_{j,t}^g} \omega_{i,t} \cdot (WageGrowth)_{i.t}}{\sum_{i \in \mathscr{J}_{j,t}^g} \omega_{i,t}},$$

Figure 2 presents the location-specific wage growths of younger and older workers. The spatial distribution differs markedly across age groups. For younger workers, higher wage growth is concentrated in high-productivity regions, particularly in urban locations in the productive South-East (London, Reading, Oxford, Cambridge). For older workers, it is consistently low

[12]Since our analysis starts in 2004 and we look at wage growth over a 3-years period, we do not have observations for 2004 and 2005 wage growth.

Figure 2: Wage growth for stayers by age

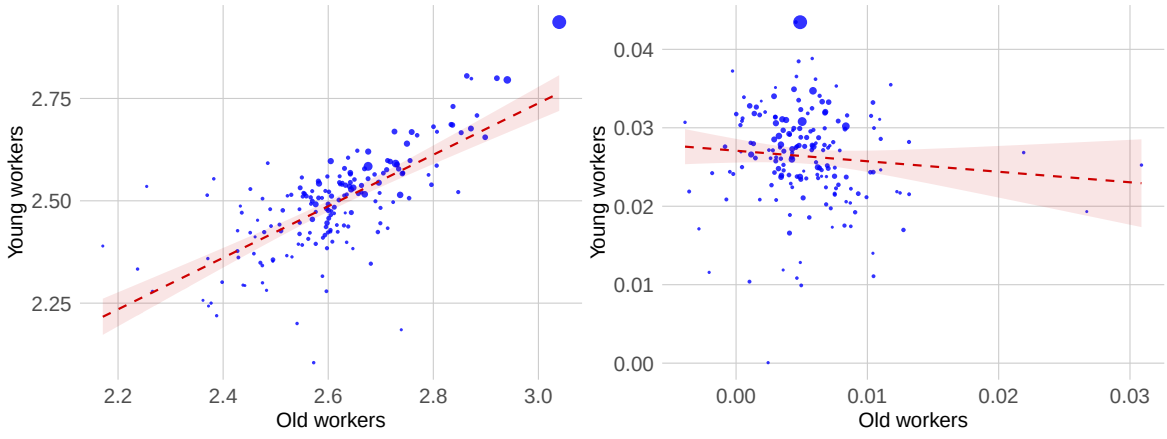


(a) Young (b) Old

Notes: The figure shows the average wage growth, expressed in annualised percentage points, for individuals who did not leave the workforce or changed location over a period of at three years ("stayers"). Panel (a) reports the figure for individuals aged 25-34, and panel (b) for individuals aged 50-64 years old. Locations with wage growth above the age-specific median are depicted in increasingly darker shades of blue, and those below in increasingly darker shades of red. Those around the age-specific median are in white.

and spatially homogeneous. Therefore, the spatial correlation of young and old wage growth is basically null, as displayed in Figure 3b.

Figure 3: Average wage and annualised wage growth

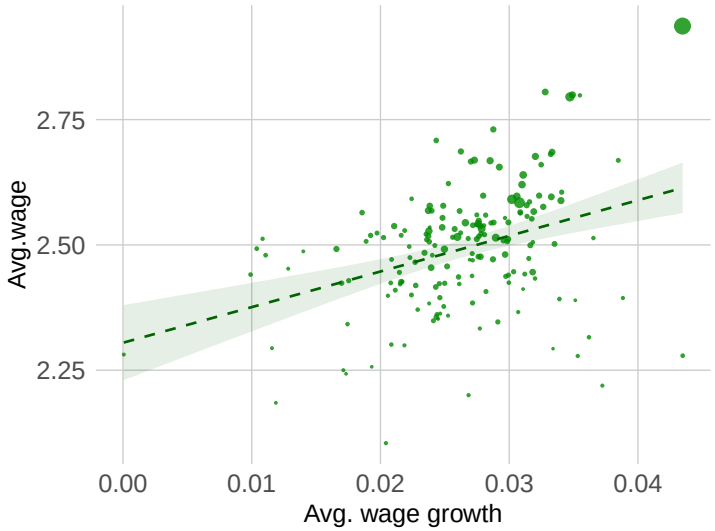


(a) Average Wage (b) Average wage growth

Notes:The figure plots the cross-sectional relationship between young and old workers' labor market outcomes across Travel-To-Work-Areas. Panel (a) shows the average log real hourly wage for young workers (aged 25–34) on the vertical axis against the average log wage for old workers (aged 50–64) on the horizontal axis. Panel (b) shows the annualised wage growth over a three year horizon for young workers on the vertical axis against wage growth for old workers on the horizontal axis. All measures are computed for "stayers"—individuals who did not leave the workforce or change location over a period of three years. The dashed lines represent fitted linear relationships, with the associated confidence interval. Data cover the period 2004–2019.

Taken together, these findings highlight the potential benefits of working in the most productive labor markets, and how they differ for young and old workers. While those labor market provides high wages for both age groups, young worker receive an additional benefit from those locations: higher wage growth, which they can spend during their longer residual work life. This spatial correlation between young's wages and wage growth is displayed in Figure 4.

Figure 4: Average wage and annualised wage growth for young workers



Notes: The figure plots the relationship between average wage levels and average wage growth for young workers (aged 25–34) across Travel-To-Work-Areas. Each point represents a TTWA. The horizontal axis shows the annualised wage growth over a three year horizon, while the vertical axis shows the average log real hourly wage. The dashed line shows the fitted linear relationship with 95% confidence interval (shaded area). All measures are computed for "stayers"—workers who remained in the same travel to work area for three years. Data cover the period 2004–2019.

2.3 Fact 3: Aging and Productivity

Like many advanced economies, England and Wales have experienced steady population aging over the past two decades, with the average age rising from 38.3 years in 2002 to 40.2 years in 2019.[13] However, aggregate statistics mask substantial spatial heterogeneity in demographic change. Since our focus is on how age composition affects labor market outcomes, we construct a measure that captures shifts in the age structure specifically within the working-age population. We define *Workforce Aging* (henceforth *Aging*) as the change in the relative representation of workers approaching retirement versus those in early career stages. Specifically, for a given location j and year t, we measure the difference between the share of older workers (aged 50–64) and the share of younger workers (aged 25–34) within the total working-age population

[13] Source: Office for National Statistics.

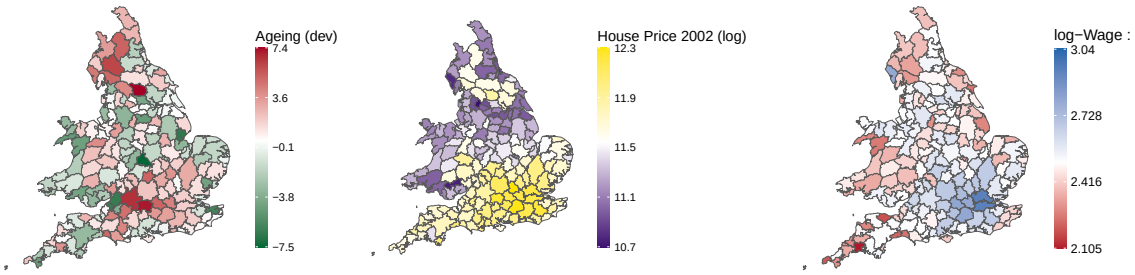
(aged 16–64). The h-year change in Aging is:

$$Aging_{j,t,h} = \Delta^h \left(\frac{N_{j,t}^{50\text{–}64} - N_{j,t}^{25\text{–}34}}{N_{j,t}^{16\text{–}64}} \right),$$

where $N_{j,t}^{g}$ denotes the population in age group g in location j and year t, and Δ^h represents the h-year difference operator. Positive values indicate an aging workforce, reflecting either an increase in the share of older workers, a decrease in the share of younger workers, or both.

At the national level, England and Wales experienced an Aging of 3.6 percentage points between 2002 and 2019, driven by the share of older workers rising from 27.3% to 30.6% while the share of younger workers declined modestly from 21.8% to 21.5%. However, this aggregate trend conceals striking spatial variation. Figure 5a displays the change in Aging between 2002 and 2019 for each TTWA, measured as deviations from the national average. The map reveals that workforce aging is far from uniform across space. The most productive areas—particularly in southern and eastern England have aged disproportionately rapidly. These locations experienced substantially larger increases in the share of older workers relative to younger ones compared to the national trend. Critically, these rapidly aging areas correspond precisely to the most expensive housing markets (Figure 5b) and to locations offering both the highest wages and steepest wage growth for young workers (Figure 5c). This spatial correlation between aging, productivity, and housing costs forms the empirical foundation for our analysis of intergenerational congestion in the labor market.

Figure 5: Ageing, House Price, and Value Added in England and Wales



(a) Ageing , 2002-2019 (b) House Price, 2002 (c) Wages (young)

Notes: The figure displays the spatial distribution of workforce aging, house prices, and wages across Travel-to-Work Areas (TTWAs) in England and Wales. Panel (a) shows the change in aging between 2002 and 2019, measured as the deviation from the national average (3.6 percentage points). Aging is defined as the change in the difference between the share of older workers (aged 50–64) and younger workers (aged 25–34) relative to the total working-age population (aged 16–64). Panel (b) shows the log house price in 2002. Panel (c) reports the average log real hourly wages for young workers (aged 25–34) who did not leave the workforce or change location over a period of three years ("stayers").

We formally test the relationship between local aging patterns, wages, wage growth, and house

prices using the following cross-sectional regression:

$$Aging_j = \beta_0 + \gamma_1 \log(\text{Wage}_j^Y) + \gamma_2 \log(\text{WageGrowth}_j^Y) + \gamma_3 \log(\text{Density}_j) + \gamma_4 \log(\text{HP}_j) + \varepsilon_j,$$

where Wage_j^Y and WageGrowth_j^Y are the TTWA-specific wage level and annualized wage growth for young workers estimated in Section 2.1, Density_j is the population density of TTWA j in 2002, and HP_j is the average house price in TTWA j in 2002.[14] We measure density and house prices at the beginning of the sample period to mitigate potential reverse causality from demographic changes affecting these variables over time.

Table 3 presents the results. Column (1) reveals a striking pattern: TTWAs offering higher wages to young workers have aged *faster*. While this may initially appear counterintuitive, it becomes interpretable when we recognize that these locations also offer high wages to older workers. This positive relationship suggests that older workers' location decisions crowd out younger workers from high-wage locations despite the strong wage incentives for the youth to locate there. These are also the locations where the young's wage growth is fastest: column (2) shows how local young wages' growth is positively correlated with ageing. Column (3) controls for both wage level and growth, showing that - given the wage level - young workers' wage growth does not have a statistically significant relationship with ageing. Again, this is to be expected since the wage level captures most of the correlation between young and old workers' location incentives.

These patterns are robust to controlling for population density, which accounts for young workers' documented preference for urban areas, and the initial house price levels in column (4). The house price coefficient is positive and highly significant: locations that were more expensive in 2002 experienced disproportionately rapid aging.[15] Importantly, including house prices does not eliminate the positive wage coefficient, suggesting that both labor market opportunities and housing are important to explain the local age composition change.

Column (5) replaces young workers' wages with old workers' wages, confirming that locations offering high labour market returns to older workers age faster. Columns (6) and (7) decompose the aging measure into the change in the share of young workers and the change in the share of old workers. The results show that aging in high-wage, high-price locations reflects two reinforcing forces rather than a single demographic shift. TTWAs with high wages for the young saw both a decline in young workers' share and an increase in older workers' share. This pattern indicates that aging in productive areas is not driven simply by uneven age-group growth, but by active reallocation: young workers are displaced from—or discouraged from

[14] Source: Land Registry price paid data.

[15] Note that we use 2002 values for density and house prices to avoid any possible reverse causality.

entering—these locations, while older workers increasingly concentrate there. This suggests a possible intergenerational congestion in spatial labor markets.

Taken together, these patterns suggest that mortgage market frictions have the potential to generate spatial misallocation: the friction contributes to systematically price younger workers out of locations where their human capital would grow fastest, while older workers—facing lower housing costs due to accumulated wealth—disproportionately occupy these productive locations despite lower returns to experience. This is a novel channel through which housing finance frictions distort not only the timing and nature of homeownership decisions across the life cycle, but also the spatial distribution of human capital with potentially significant aggregate consequences. Understanding the magnitude of these spatial distortions and designing policies to mitigate them is essential for both efficiency and equity.

Table 3: Ageing, Wages and Return to Experience

Dependent variable:	Ageing					Change in share young	Change in share old
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(Intercept)	-11.87***	0.8497	-16.07**	-29.41***	-27.18***	33.11***	3.696
	(3.382)	(0.8146)	(6.514)	(9.291)	(8.168)	(5.008)	(4.787)
(log) Wage young	5.981***		8.415***	6.472***		-3.306***	3.166**
	(1.373)		(3.132)	(1.721)		(0.9496)	(1.444)
Wage growth young		95.66**	-71.04	-42.29		17.75	-24.54
		(40.89)	(54.51)	(43.07)		(22.94)	(27.57)
(log) Density 2002				-0.8678***	-0.8675***	0.1586	-0.7092***
				(0.2328)	(0.1906)	(0.1329)	(0.1707)
(log) House Price 2002				2.501***	2.461***	-2.356***	0.1452
				(0.6863)	(0.7228)	(0.4411)	(0.3966)
(log) Wage old					5.143***		
					(1.790)		
Wage growth old					-37.13		
					(52.49)		
Observations	173	173	173	173	173	173	173
R^2	0.131	0.058	0.141	0.328	0.330	0.568	0.163

Notes: OLS estimates with Conley (75km) standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Dependent variables: "Ageing" = Change in share of old vs. young (2002–2019); "Change in share young" = Change in share of workers aged 25–34 (2002–2019); "Change in share old" = Change in share of workers aged 50–64 (2002–2019). All dependent variables are expressed in percentage points. Wage variables and density are measured in logs. House prices are measured in 2002. Sample includes 173 TTWAs in England and Wales. All wage and wage growth measures are computed for workers who remain in the same TTWA for at least three years ("stayers"), as described in Section 2.1.

3 A Tractable Model of Intergenerational Congestion

In this section, we introduce a tractable two-location overlapping generations model to illustrate the macroeconomic consequences of intergenerational competition for space. We show that the demographic composition of cities is inefficient whenever older workers face lower housing costs than the young.[16] Moreover, spatially correlated labor market incentives across

[16]Importantly, we do not need such cost differences to be spatially correlated with local productivity or learning, as long as housing demand is non-homothetic in income. Which, indeed, is empirically the case.

age groups exacerbate the inefficiency. After outlining the model, we present analytical results and emphasize the key economic intuition behind our findings.

3.1 Environment

Time is discrete, $t = 1, 2, 3, \ldots$, and the economy continues indefinitely. In each period, a new generation of young workers of measure 1 is born. These individuals live for two periods: they are *young* (superscript Y) in period t and become *old* (superscript O) in period $t+1$, after which they die. This OLG structure follows Samuelson (1958) and Diamond (1965a).

Agents choose where to live in each period of their lives. There are two locations: the *city* (C) and the *rural area* (R). Each location $j \in \{C, R\}$ is characterized by productivity Z_j, location-specific human capital returns e_j, age-specific amenities μ_j^g for $g \in \{Y, O\}$, and fixed housing supply n_j. All of these features are constant over time. Importantly, all young individuals are born with identical skill level s^Y, but their skill as old workers, s^O, depends on the location they chose when young.

Following Rosen (1979) and Roback (1982), location choice is discrete and driven by idiosyncratic preferences. Each worker, indexed by i, draws idiosyncratic utility shocks $\varepsilon_{i,j,t}$ from a location-, age-, and skill-specific distribution $\mathscr{F}_{j,t}^{g,s}$, where g denotes age and s denotes skill. Let $\varepsilon_{i,t} = \{\varepsilon_{i,C,t}, \varepsilon_{i,R,t}\}$ be the vector of utility shocks in period t. Workers choose their location in each period: j^Y when young and j^O when old, where $j^g \in \{C, R\}$.

Living in location j incurs a housing cost of $\psi^g P_{j,t}$, where ψ^g is age-specific. We normalize $\psi^Y = 1$, so that ψ^O captures the relative housing cost for old workers. We assume $\psi^O = (1 - \psi)$ with $0 < \psi < 1$, consistent with the empirical evidence in Section 2.1. Thus, ψ measures the strength of the housing market friction that makes locations more expensive for young relative to old workers. In this tractable model, we treat ψ as exogenous and focus on its implications for spatial sorting rather than its micro-foundations.

Production Each location hosts a continuum of competitive firms producing a tradeable good using labor as the only input:

$$Y_{j,t} = Z_j L_{j,t},$$

where Z_j is location-specific productivity and $L_{j,t}$ is total labor input measured in efficiency units. Firms maximize profits, $\pi_{j,t} = Y_{j,t} - w_{j,t} L_{j,t}$, and free entry drives profits to zero in equilibrium. This implies wages equal productivity: $w_{j,t} = Z_j$. Since profits are zero, firm ownership is irrelevant.

Initial condition and information set The economy begins with an initial generation of old workers with exogenous skill level s_1^O. In each period of their lives, agents observe the time-invariant characteristics of both locations and their current-period idiosyncratic preference shocks before making their location choice.

3.2 Workers' problem

We now characterize the workers' optimization problem.

Preferences An individual of age $g \in \{Y, O\}$ has intratemporal utility given by:

$$U_i(c_t^g, j) = c_t^g + \mu_j^g + \varepsilon_{i,j,t},$$

where c_t^g is consumption of the tradeable good, μ_j^g denotes age-specific amenities in location j, and $\varepsilon_{i,j,t}$ captures idiosyncratic utility from living in location j. The utility function is linear in consumption, implying risk neutrality. We relax this assumption in the quantitative model of Section 4.

Budget constraint When living in location j, a worker i of age g pays the housing cost $\psi^g P_{j,t}$ and earns labor income $h_{i,j,t}(s_i)$, where s_i denotes their skill level.[17] Labor income is determined by the local wage rate $w_{j,t}$ and individual skill: $h_{i,j,t}(s_i) = w_{j,t} s_i$.

Regardless of their location, workers receive a share of the profits of the national housing sector, denoted as $\tau_{i,t}$. This revenue is distributed uniformly to all the individuals alive in a lump-sum manner, so that $\tau_{i,t} = \tau_t$, $\forall i$.[18] Additionally, workers have access to a risk-free asset b_{t+1} with return r_t. We assume $r_t = 0$ for all t and that young workers do not discount future utility.

The budget constraint of individual i of age $g \in \{Y, O\}$ in location j at time t is:

$$c_t^g + \psi^g P_{j,t} + b_{t+1}^g \leq w_{j,t} s_i + \tau_t + b_t^g, \quad \text{with } g \in \{Y, O\}, \quad c_t^g \geq 0,$$

where $b_t^Y = 0$ (young workers are born with no assets) and $b_{t+1}^O = 0$ (old workers do not save since they die next period and cannot borrow as they would not repay).

We now introduce key assumptions that ensure tractability and interior solutions:

Assumption 1.

[17] Recall that there are three possible skill levels: s^Y for the young, and $s^O(j^Y = C)$ and $s^O(j^Y = R)$ for the old.

[18] This assumption ensures that housing rents remain within the economy, which simplifies the social planner's problem.

S.1 *(All locations are affordable) Exogenous parameters satisfy* $\tau_t + h_{i,j,t}(s_i) \geq \psi^g P_{j,t}$ *for all agents* i *of age* g *and skill* s*, all locations* j*, and all periods* t*.*

S.2 *(Difference is uniformly distributed) For each individual* i*, in any period and for any skill:*[19]

$$\varepsilon_{i,j} - \varepsilon_{i,\bar{j}} \sim U(-\sigma, \sigma), \quad j \neq \bar{j}.$$

S.3 *(Mixing):*

$$\sigma > |\Delta\mu^Y| + |\Delta\mu^O| + 2|\Delta Z(1 + e + e_C)| + |Z_R e_C|.$$

S.4 *(Centering) For any skill,* $\varepsilon_{i,j}$ *and* $\varepsilon_{i,\bar{j}}$ *satisfy:*

$$E_t(V^O_{t+1}(s^O(j^Y), \varepsilon_{t+1})) = \sum_{j'} \phi^O_{j',t+1}(s^O(j^Y)) \left(\mu^O_{j'} + h_{i,j',t+1}(s^O(j^Y)) - \psi^O P_{j',t+1}\right).$$

S.5 *(Equal means)* $\varepsilon_{i,j}$ *and* $\varepsilon_{i,\bar{j}}$ *have identical unconditional means:*

$$E_{t-1}(\varepsilon_{i,j,t}) = E_{t-1}(\varepsilon_{i,\bar{j},t}), \ \forall t, i.$$

Assumption S.1 ensures that all workers can afford housing in any location based solely on their labor income and housing profit share. Thus, all locations are feasible for all agents. Combined with utility linearity and the assumption that r_t equals the discount rate, Assumption S.1 implies that young workers' saving decisions do not affect their optimal location choice. Given the optimal location choice, any allocation $\{b^Y_{t+1}, c^Y_t, c^O_{t+1}\}$ satisfying the intertemporal budget constraint yields identical lifetime utility. Therefore, the free-borrowing equilibrium coincides with the no-borrowing equilibrium. For expositional simplicity, we impose $b^Y_{t+1} = 0$ for all t, but emphasize that the optimal spatial allocation is not driven by borrowing constraints. Assumption S.2 follows **?**, while Assumption S.3 guarantees interior solutions with both young and old workers in each location. Assumptions S.4 and S.5 are normalizations that ensure analytical tractability.[20]

At the optimum, the budget constraint binds, so per-period utility for an individual of age g

[19] See Moretti (2011).

[20] Assumption S.4 ensures that the expected continuation value for young workers does not depend on the relative magnitude of idiosyncratic versus systematic utility components when old. This parallels the standard practice in discrete choice models with Gumbel shocks, where expected utility is adjusted by $-\gamma\nu$ (with γ the Euler-Mascheroni constant and ν the scale parameter). Since higher realizations of ε influence location choice, the expected value of ε conditional on being chosen exceeds its unconditional mean. Assumption S.4 introduces an offset—potentially skill-specific—to neutralize this selection premium. The model allows a closed-form solution for the steady state without this assumption, but it is complex enough to provide little economic intuition.

depends only on location j and skill s_i:

$$U_i^g(s_i, j) = \mu_j^g + w_{j,t} s_i + \tau_t - \psi^g P_{j,t} + \varepsilon_{i,j,t}.$$

We now characterize the problem for each age group. Suppressing the individual index i, the old workers' utility maximization problem is:

$$V^O(s^O(j^Y), \varepsilon_t) = \max_{j \in \{C,R\}} \left\{ \mu_j^O + w_{j,t} s^O(j^Y) + \tau_t - (1-\psi) P_{j,t} + \varepsilon_{j,t} \right\},$$

where j^O denotes the optimal location choice. Old workers solve a static problem, choosing location j^O to maximize current utility conditional on their skill $s^O(j^Y)$ and preference shocks ε_t.

Young workers solve a dynamic location choice problem:

$$\begin{aligned} V^Y(s^Y, \varepsilon_t) &= \max_{j \in \{C,R\}} \left\{ \mu_j^Y + w_{j,t} s^Y + \tau_t - P_{j,t} + \varepsilon_{j,t} + \mathbb{E}_t[V^O(s^O(j), \varepsilon_{t+1})] \right\}, \\ \text{s.t.} \quad & s^O(j) = s^Y(1 + e + e_j), \end{aligned}$$

where the expectation is taken over the distribution of next-period preference shocks ε_{t+1}. Young workers account for how their location choice affects both current income and future skill accumulation. Note that housing costs differ by age: $P_{j,t}$ for the young versus $(1-\psi)P_{j,t}$ for the old.

In Appendix B.1, we derive the location choice probabilities: $\phi_{j,t}^O(s^O(j^Y))$ for old workers with skill $s^O(j^Y)$ and $\phi_{j,t}^Y(s^Y)$ for young workers.

3.3 Equilibrium and Intergenerational Congestion

To characterize the equilibrium, we introduce simplifying normalizations without loss of generality. We assume equal housing supply across locations, $n_C = n_R = 1$, normalize the initial skill level to $s^Y = 1$, and set the rural price to $P_R = 0$. Since housing supply is fixed, only the difference price $P_C - P_R$ is determined in equilibrium, making the rural price normalization innocuous. Under this normalization, the city price P_C represents the additional cost of living in the city relative to the rural area. Finally, we assume $e_R = 0$, so that only the city provides location-specific skill accumulation.

Before analyzing the sequential market equilibrium, we define the steady-state equilibrium as an allocation and price system satisfying all equilibrium conditions that remains constant over time. We denote steady-state variables with a tilde (e.g., $\tilde{P}_C$). A formal definition is provided in

Appendix B.2.

Proposition 1 (Steady-state equilibrium)**.** *Assume Assumptions S.1–S.4 hold, and that* $n_C = n_R = 1$, $P_R = e_R = 0$, *and* $s^Y = 1$. *Then, the system of equations defining the sequential market equilibrium (equations (17)) admits a unique steady state* $\tilde{\phi}_C^Y$ *given by:*

$$\tilde{\phi}_C^Y = \frac{1}{2} + \frac{\overbrace{(1-\psi)\Delta\mu^Y - \Delta\mu^O}^{\textit{Net amenity pull}}}{(4-2\psi)\sigma + (3-2\psi)\Delta Z\, e_C + (1-\psi)\frac{(\Delta Z e_C)^2}{\sigma}} + \underbrace{\frac{\overbrace{(1-\psi)Z_R e_C - e\Delta Z}^{\substack{\textit{Human capital considerations} \\ \textit{(scaled by housing friction)}}}}{(4-2\psi)\sigma + (3-2\psi)\Delta Z\, e_C + (1-\psi)\frac{(\Delta Z e_C)^2}{\sigma}}}_{\textit{Denumerator: Spatial concentration of wages and returns to experience}} - \frac{\overbrace{\psi\Delta Z(1+\frac{e_C}{2})}^{\substack{\textit{Housing cost friction:} \\ \textit{inaccessible wage premium}}}}{(4-2\psi)\sigma + (3-2\psi)\Delta Z\, e_C + (1-\psi)\frac{(\Delta Z e_C)^2}{\sigma}}. \tag{1}$$

The market allocation in equation (1) reflects three distinct forces shaping the equilibrium share of young workers in cities. First, *amenity considerations.* The young respond to their net preference for the city $\Delta\mu^Y$, but their responsiveness is scaled by $(1-\psi)$, reflecting their higher housing costs. This amenity pull is further dampened by old workers' own amenity incentives ($\Delta\mu^O$), which increase competition for city space. Second, *human capital accumulation.* Young workers are pulled to the city by the promise of future skill gains e_C. City-specific experience has option value ($Z_R e_C$): skills acquired in the city remain productive even if workers later relocate to rural areas. Again, this force is dampened by the higher housing costs faced by the young (ψ), and by the competition for space coming from the fact that all old individuals are at least e more skilled than the young. However, cities offer high wages when workers are old regardless of their early-life location ($e\Delta Z$), since general experience e combines with the city productivity advantage ΔZ. Higher general experience e increases the competition for city space arising from the old. Third, *housing cost distortions.* The term $-\psi\Delta Z(1+\frac{e_C}{2})$ captures the distortionary effect of age differences in housing costs. By paying higher housing costs, the young workers are crowded out by the old, in a way proportional to the productivity advantage ΔZ (identically skilled old can afford the city, as they pay less than the young) and the urban wage premium of urban experience $\Delta Z e_C$ (the more the old with city experience are higher-skilled than the young, the more they are willing to remain in the city the next period). When $\psi = 0$, this distortion vanishes and young workers are more likely to locate in the productive

location.

This discussion highlights how this model features what we call *intergenerational congestion*: by demanding more city space for themselves, the old crowd out the young. Conversely, the larger the current and *expected* gains from working in the city, the more the young are willing to move to the city, crowding out older, skilled workers. Most importantly, the amount of congestion caused by the old onto the young is *increasing* in the complementarity between city skills and city productivity, ΔZe_C. The larger the complementarity premium, the "older" the city in equilibrium, especially in the presence of housing frictions.

Proposition 3 (Appendix B.5) characterizes the sequential market equilibrium and establishes three key properties: the steady state is globally stable, convergence follows a cyclical path, and the equilibrium share of young workers in the city remains interior (i.e., $\tilde{\phi}_C^Y \in (0,1)$) for all admissible initial conditions.

3.4 Human Capital Misallocation

In this section, we assess whether the market equilibrium allocation is efficient. We show that intergenerational congestion—which discourages young workers from locating in the city and accumulating human capital—generates welfare losses that can be corrected by a constrained-efficient planner.

We formally establish that a social planner who can assign consumption bundles to agents of each skill level, age, and location, while allowing free spatial sorting (following Lhuillier (2024)), implements an allocation with more young workers in the city than the market equilibrium whenever $\psi > 0$. The inefficiency arises because in that case the social value and market price of city space are misaligned. A planner can realign private incentives with social costs and benefits through age- and location-specific transfers (equivalent to differential taxes and subsidies), thereby correcting the spatial misallocation.

A utilitarian social planner maximizes the present discounted value of aggregate utility across all current and future generations. The planner observes all fundamentals—productivity, amenities, and the distribution of idiosyncratic preferences—but does not observe individual preference realizations. In this first-best analysis, the planner can condition consumption allocations on observables: age g, location j, and skill s. Thus, the planner chooses consumption schedules $\{c_j^g(s)\}_{j\in\{C,R\},g\in\{Y,O\}}$ for each type of agent. Given these consumption allocations, agents freely choose their location to maximize utility based on their realized idiosyncratic preferences.

The timing is as follows: the planner commits to consumption schedules *before* idiosyncratic shocks are realized, then agents observe their preference shocks and choose locations opti-

mally. Under this timing, the planner maximizes the expected utility of each generation, where the expectation is taken over the distribution of idiosyncratic preferences. While this timing assumption does not affect the planner's optimal allocation, it is crucial for establishing that Pareto improvements exist relative to the market equilibrium.

Formally, the planner problem can be written recursively as:

$$\begin{aligned}
\hat{V}(\mathbb{S}) = \max_{\{c_j^g(s)\}_{j\in\{C,R\},g\{Y,O\}}} \int \mathbb{E}[U_i(s_i,j_i)]\,di + \mathbb{E}\left[\hat{V}\left(\mathbb{S}', \{\varepsilon_i'\}_{i\in[0,2]}\right)\right] \quad \text{s.t.} \qquad (2)\\
n_j = \int \mathbb{1}\{j_i = j\}di \quad j \in \{C,R\},\\
\int Z_{j_i} s_i di = \int_{i:g_i=Y} c_{j_i}^Y(s_i)di + \int_{i:g_i=O} c_{j_i}^O(s_i)di\\
\mathbb{S} \equiv \{s_i\}_{i\in[0,2]},\\
\mathbb{S}' = \begin{cases} s^Y \quad \text{with mass 1} \\ (1+e+e_C)s^Y \quad \text{with mass } \int \mathbb{1}\{j_i = C\}\,di \\ (1+e+e_R)s^Y \quad \text{with mass } \int \mathbb{1}\{j_i = R\}\,di \end{cases}
\end{aligned}$$

where $\mathbb{S}$ denotes the skill distribution in the economy. Note that the planner directly allocates agents to locations subject to space constraints $n_C = n_R = 1$, bypassing the housing market entirely. Since all revenues generated by the housing sector in the market economy are distributed lump-sum to the agents, the total resources available for consumption are the same for both the planner and the market economy, for any given allocation. Therefore, the absence of a housing market does not impact overall welfare.

Proposition 2 characterizes the planner's optimal allocation in terms of the steady-state share of young workers in the city, $\hat{\phi}^Y$.

Proposition 2. *Assume Assumptions S.1, S.2, S.3, S.4, and S.5 hold. Also, assume that* $n_C = n_R = 1$, $P_R = e_R = 0$, *and* $s^Y = 1$.

P.5 *The steady-state solution of the planner problem, expressed as share of young that are allocated to the city, is:*

$$\hat{\phi}^Y = \tilde{\phi}^Y + \psi\left[\underbrace{\frac{\Delta\mu^y + \Delta Z - 2\sigma\left(\tilde{\phi}^Y - \frac{1}{2}\right)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}}}_{\substack{\text{Static sorting wedge:}\\ \text{Direct period utility gain from city location}}} + e_c \underbrace{\frac{Z_R + \Delta Z\left(\frac{1}{2} - \left(2 + \frac{\Delta Z e_c}{\sigma}\right)\left(\tilde{\phi}^Y - \frac{1}{2}\right)\right)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}}}_{\substack{\text{Dynamic human capital wedge:}\\ \text{Option value of experienc utility gain}}}\right].$$

where $\tilde{\phi}^Y$ *is the steady-state market share of young in the city from Proposition P.1.*

P.6 *If* $\psi > 0$, *and* $Z_R e_C > \Delta Z\left(e + e_C\left(1 + \frac{\Delta Z e_C}{2\sigma}\right)\right)$, *then the aggregate output of the economy*

is always strictly larger under the planner's allocation, relative to the market's allocation.

P.7 *Consider a planner facing a Ramsey problem. The only policy instrument at disposal of the planner is a set of age-location-specific tax rates* $\{\tau^h_{a,j}\}_{j\in\{R,C\},a\in\{Y,O\}}$ *on housing expenditure* P_{jt}*, whose revenue is then distributed lump-sum. The planner's constrained second-best corresponds to the planner's first-best from Proposition P.7, and is implmented with the tax set*

$$\tau_{a,j} = \begin{cases} \frac{\psi}{1-\psi} & \text{if } a = O, \forall j \\ 0 & \text{if } a = Y, \forall j \end{cases}.$$

See Appendix B.7 for the proof.

Proposition 2 establishes that the market equilibrium is efficient if and only if $\psi = 0$, that is, when young and old workers face identical housing costs. When $\psi \neq 0$, the market features dynamic misallocation. In the empirically relevant case for the UK where $\psi > 0$ (young workers face higher housing costs) and cities are both more productive and offer higher returns to experience, the market equilibrium has too many old workers and too few young workers in the city, resulting in an inefficiently low aggregate human capital in equilibrium.

The misallocation arises because workers do not internalize how their location choices affect the age composition of the city through equilibrium prices. Since old workers pay $(1-\psi)P_j$ rather than the full social cost of occupying urban space, market incentives diverge from social efficiency. Intergenerational congestion generates inefficiency whenever $\psi \neq 0$ and locations differ in productivity, returns to experience, *or* amenities.

This inefficiency operates through two distinct channels. First, the *static sorting wedge* captures the immediate utility loss from misallocation: old workers with lower idiosyncratic preferences for the city displace young workers with higher city-specific preferences, reducing the quality of sorting on amenity values. This wedge reflects forgone consumption (when output is lower) and forgone utility from better matches between workers and locations. Second, the *dynamic human capital wedge* captures the long-run cost of insufficient skill accumulation: when young workers are priced out of high-learning cities, they miss critical learning opportunities (when $e_C > 0$) and accumulate too little human capital, reducing both their future productivity and total lifetime earnings. This dynamic loss persists across generations, as lower skill accumulation today reduces the economy's productive capacity in all future periods.

Proposition P.6 establishes conditions under which the planner's allocation generates strictly higher aggregate output. This occurs when the option value of city-specific experience is sufficiently large—that is, when skills acquired in the city remain productive if workers later relocate

to rural areas ($Z_R e_C$ is large) relative to the wage premium from using general experience in the city ($\Delta Z e$). In this case, the market's excessive concentration of old workers in cities is particularly costly, as it prevents young workers from acquiring skills that would be valuable regardless of their future location choices.[21]

Finally, Proposition P.7 establishes that the planner's first-best can be achieved also in a Ramsey problem where the planner has at its disposal only age-place-based taxes. The optimal tax rates cancel all differences in house prices between age groups, which - in this simplied model - corresponds to setting the price of housing equal to its marginal (social) cost.

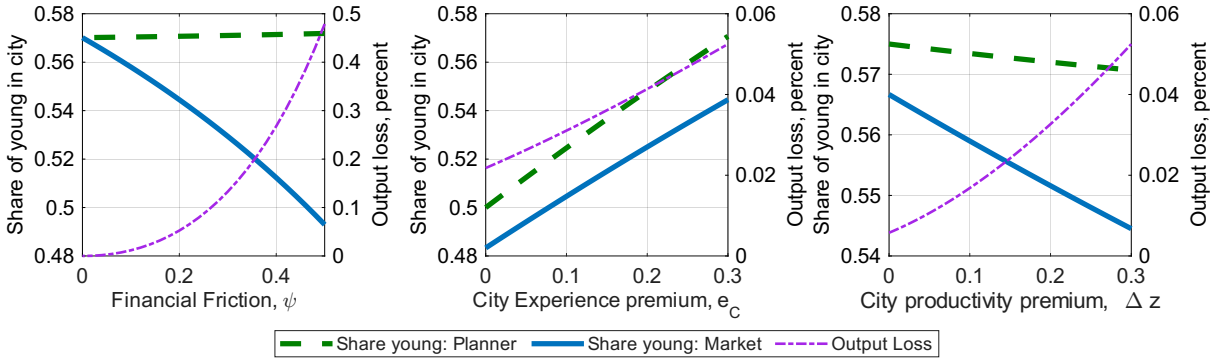
Figure 6 illustrates the steady-state allocation of young workers and the associated output losses as a function of key model parameters. Each panel plots the share of young workers in the city under the market equilibrium (blue solid line, left axis) and the planner's allocation (green dashed line, left axis), along with the percentage output loss from the market misallocation (purple dotted line, right axis).

The left panel shows comparative statics with respect to the housing friction ψ. As ψ increases (young workers face progressively higher housing costs relative to old workers), the market allocation features fewer young workers in the city, while the planner's optimal allocation remains relatively stable. Consequently, the wedge between the two allocations widens, and output losses increase monotonically. Notably, when $\psi = 0$ (no housing cost differential), the market and planner allocations coincide, confirming the efficiency of the market equilibrium absent housing frictions. The middle panel examines the role of city-specific experience returns e_C. Higher returns to city experience increase both the market and planner allocations of young workers to the city, as skill accumulation becomes more valuable. However, the planner allocates consistently more young workers than the market. The output loss increases with e_C This reflects the growing importance of the *dynamic human capital wedge*: when city experience is highly valuable, the market's failure to allocate sufficient young workers to skill-building locations becomes increasingly costly. The right panel shows comparative statics with respect to the city productivity premium ΔZ. Interestingly, higher city productivity decreases the market share of young workers while increasing the planner's allocation, causing the wedge between them to widen substantially. The market allocation declines because higher ΔZ attracts more old workers to the city (who value current wages but not skill accumulation), intensifying intergenerational congestion and crowding out the young. Meanwhile, the planner recognizes the greater social value of allocating young workers to high-productivity cities for skill development, highlighting how spatial concentration of both productivity and learning opportunities exacerbates misallocation. Across all panels, the key insight is that the market systematically

[21] If we interpret idiosyncratic utility as idiosyncratic productivity in home production across locations, then output losses would coincide exactly with welfare losses.

under-allocates young workers to productive, high-learning cities whenever $\psi > 0$, and this inefficiency intensifies as cities become more attractive (either through higher productivity or greater skill accumulation opportunities).

Figure 6: Steady State Comparative Static



Notes: the figure compares the share of young in the city (ϕ^Y) at the market steady state (blue line) and the planner steady state (dashed green line). The purple line show the output gap between the planner and the market allocation, in percentage points (right-hand scale). The left-hand panel shows how the allocation and output gap vary along with e_C, the experience premium of the city, assuming a baseline $\Delta Z = 0.3$. The right-hand panel shows how the allocation and output gap vary with ΔZ, the productivity gap between City and Rural, assuming a baseline $e_C = 0.3$. Both graphs assume $\sigma = 1$, $Z_R = 1$, $e = 0$, and $\Delta\mu^Y = \Delta\mu^O = 0$.

Appendix B.8 demonstrates that, under mild conditions on young workers' income, there exists a Pareto-improving allocation that attains the planner's optimum while making no generation worse off. The mechanism works as follows: young workers in period t make transfers to old workers, inducing the old to relocate from cities to rural areas despite facing lower wages. These transfers compensate old workers for their ex-ante expected income loss (measured before idiosyncratic preference shocks are realized). In period $t+1$, when these young workers become old, they receive transfers from the next generation of young workers, ensuring they are fully compensated for their earlier contribution. The policy creates a Pareto improvement through intergenerational redistribution. By incentivizing old workers to vacate city space, young workers gain access to high-productivity, high-learning locations, accumulating more human capital and earning higher lifetime incomes. This increased productivity generates sufficient resources to compensate both the initial old generation (who relocate) and all future generations (through the transfer chain). Since all agents evaluate the policy before observing their idiosyncratic preferences, and the expected utility gain from better spatial sorting exceeds the transfer costs, all generations are weakly better off in expectation, with most strictly better off.

4 Quantitative Model

We extend the tractable model of Section 3 in several dimensions to provide convincing and realistic quantitative estimates of the magnitude of intergenerational congestion and its macroeconomic effects. The quantitative framework incorporates key features from recent advances in spatial economics and household finance: we adopt a multi-location spatial structure with endogenous migration and skill accumulation following Greaney et al. (2025) and Lhuillier (2024), and introduce rich life-cycle heterogeneity in housing tenure decisions and borrowing constraints drawing on Kaplan et al. (2020). Critically, we endogenize the housing cost differential between young and old workers by modeling homeownership choices, optimal deposit decisions, and leverage-dependent mortgage rates, thereby microfounding the reduced-form friction ψ from the theoretical model. This extension allows us to quantify welfare losses, evaluate the distributional consequences of intergenerational congestion, and assess targeted policy interventions.

4.1 Fundamentals

4.1.1 Timing

Overlapping generations populate the economy. Age is indexed by $a \in [0, \overline{A}]$. Agents work between ages $[a, A_{\text{ret}})$, are retired between ages $[A_{\text{ret}}, \overline{A})$, and die at age $\overline{A}$, where A_{ret} and $\overline{A}$ are integers. Following Greaney et al. (2025), we separate individuals' lives into "shock ages", denoted by $\mathscr{A}^s = \{0, 1, \ldots, \overline{A} - 1\}$, and the remainder of the lifespan. At shock ages, individuals receive productivity and location preference shocks and are allowed to re-optimize their location and housing choices. Between shock ages, individuals face no shocks and solve a standard consumption-saving problem conditional on their existing location and housing stock. Time is continuous and calendar time is indexed by t.

4.1.2 Locations

The economy consists of J locations, indexed by $j = 1, \ldots, J$. Each location is characterized by three fundamental attributes: innate productivity Z_j, housing sector productivity Z_j^H, and amenities μ_j. The innate productivity Z_j captures characteristics of a location and its firm environment that are independent of the spatial allocation of workers, such as the concentration of high value-added firms, quality of infrastructure, and institutional factors. We take the spatial distribution of firms as exogenous, abstracting from firm location decisions. The housing sector productivity Z_j^H reflects differences in land availability, zoning regulations, construction costs, and geographic constraints that affect the ease of housing supply expansion. Locations with higher Z_j^H can produce more housing from a given amount of construction input. The amenity

function μ_j captures location-specific consumption value in urban amenities (restaurants, culture, nightlife) versus family-oriented amenities (schools, parks, safety).[22]

4.2 Households

4.2.1 Preferences

Households derive utility from a numeraire consumption good c and housing services $\mathbf{h}$ according to the following CRRA utility function with intertemporal elasticity of substitution θ:

$$U(c,\mathbf{h},j) = \mu_j \frac{\left(c^{\sigma}\mathbf{h}^{(1-\sigma)}\right)^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}}.$$

Here, σ represents the degree of non-homotheticity of the demand functions. For $\sigma \to 0$, the utility function converges to Cobb-Douglas with an expenditure share η on housing.

At death, individuals receive utility from bequeathing liquid wealth b according to a warm glow bequest motive:

$$v(b) = \theta_0 \frac{(\theta_1 + b)^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}}.$$

The parameter θ_0 governs the strength of the bequest motive, while θ_1 ensures positive marginal utility even at zero wealth. Future utility is discounted at rate ρ.

4.2.2 Earnings

Working-age individuals supply one unit of labor inelastically. Individuals differ by skill level s, which follows an AR(1) process with age-specific drift $\gamma_0(a)$ and location-specific drift $e_j(s_i^a, a)$:

$$s_i^{a+1} = \theta_s s_i^a + \gamma_0(a) + e_j(s_i^a, a) + \zeta_i, \text{ with } \zeta_i \sim N(0, \sigma_s^2(a)),$$

where θ_s governs skill persistence, $\gamma_0(a)$ captures age-specific skill growth common across locations, and ζ_i is an idiosyncratic shock. The location-specific drift $e_j(s_i^a, a)$ captures how skills evolve differently depending on the location j where individuals have worked since the last shock age, reflecting differences in learning opportunities across labor markets.

[22] Both Z_j and Z_j^H could be generalized to depend on endogenous variables, for example, to capture agglomeration economies or congestion externalities. We abstract from these extensions to focus on the role of intergenerational housing market frictions, though incorporating such features would be straightforward.

We assume that the learning process e_j depends on local average skill $\bar{s}_j$, local innate productivity Z_j, and individual characteristics (age and skill). Following Lhuillier (2024), we allow for potential complementarity or substitutability between own and local skills. We extend their framework by incorporating a direct role for local innate productivity Z_j, capturing how learning can be faster in productive locations not only through skilled co-workers but also through firm-specific factors such as organizational practices, institutional quality, access to training resources, and superior technology. The learning function takes the form:

$$e_j(s_i^a, a) = \gamma_1(a)\ln(Z_j) + \gamma_2(a)\ln(\bar{s}_j) + \gamma_3(a)\ln(Z_j)\ln(s_i^a) + \gamma_4(a)\ln(\bar{s}_j)\ln(s_i^a). \tag{3}$$

The parameters $\gamma_1(a)$ and $\gamma_2(a)$ capture direct effects of location productivity and peer skills on learning, while $\gamma_3(a)$ and $\gamma_4(a)$ capture complementarities between individual skill and location characteristics. All parameters are allowed to vary with age to reflect life-cycle differences in learning capacity. In section C.7 we discuss the identification and estimation of this specification.

Individual earnings equal $\exp(s_i^a)w_{jt}$, where w_{jt} is the wage rate in location j at time t. Retirees receive a fixed pension benefit b_{ret}. Labor income is taxed at rate τ_y, while pensions are provided net of taxation. Thus, an individual i of age a, living in location j with skill s_i^a has earnings:

$$y_t(j, s_i^a, a) = \mathbf{1}[a < A_{\text{ret}}](1-\tau_y)\exp(s_i^a)w_{jt} + \mathbf{1}[a \geq A_{\text{ret}}]b_{\text{ret}}.$$

4.2.3 Mobility

Individuals are only allowed to change location at shock ages. Immediately after observing their productivity shock ζ_i, they draw idiosyncratic location preference shocks $\{\varepsilon_{i,j}\}_{j\in J}$ from independent Gumbel distributions with scale parameter ν. After observing these preference realizations, individuals choose their residence location to maximize expected lifetime utility. Moving from location j to location j' incurs a cost $\kappa_{j,j'}(a)$ that may depend on age and the distance between locations. Individuals work in the same location where they reside, abstracting from commuting decisions. Consistent with this assumption, our geographical units are Travel-to-Work Areas, designed to capture self-contained local labor markets.

4.2.4 Housing

Housing services can be obtained through renting or owning. Rental unit sizes h^r are restricted to the discrete set $\mathscr{H}^r$, and owner-occupied housing sizes h are restricted to the discrete set $\mathscr{H}^o$, including $h = 0$. We allow ownership to yield different utility than renting the same

physical housing unit, reflecting differences in security of tenure, customization possibilities, and psychological benefits of ownership. The total housing services enjoyed by an individual are:

$$\mathbf{h}(h, h^r) = \chi h + h^r,$$

where $\chi \geq 1$ captures the utility premium from ownership. We assume households cannot simultaneously rent and own, so that $h^r > 0 \implies h = 0$.

Housing depreciates at rate δ, which owners must fully offset through maintenance expenditures to preserve the property's value. Renters can freely adjust their rental unit size at any point in time, while owner-occupied housing size can only be adjusted at shock ages. Any adjustment in owner-occupied housing—whether in size or location—requires selling the existing property and potentially purchasing a new one. Selling a property incurs a transaction cost equal to a fraction Ψ of the housing value.

4.2.5 Budget Constraints

Households have access to a liquid, risk-free bond that serves as the sole savings vehicle. Negative bond positions (borrowing) are permitted only when secured by housing collateral. We impose two borrowing constraints reflecting institutional features of mortgage markets. First, a loan-to-value constraint limits borrowing to a fraction M_p of housing wealth. Second, an income multiplier constraint limits borrowing to a multiple M_y of current income:

$$\begin{aligned} b &\geq -M_p p_{jt} h, \\ b &\geq -M_y y_t(j, s, a). \end{aligned}$$

When either constraint binds, further borrowing is prohibited, though individuals are not required to immediately delever. Formally, wealth accumulation must satisfy:

$$\dot{b} \geq 0 \text{ if } (b \leq -M_p p_{jt} h) \vee (b \leq -M_y y_t(j, s, a)).$$

The bond is priced in units of the numeraire consumption good. The *net* interest rate on asset position b, denoted $q_t(b, j, h)$, consists of a net risk-free rate q_t plus a component that depends on the asset and liability composition of the individual's portfolio, capturing default risk in reduced form (Kaplan et al., 2020):

$$q_t(b, j, h) = q_t + \mathbf{1}[b < 0] f\left(b, p_{jt} h\right),$$

where $f(\cdot)$ is an increasing function of the individual's leverage ratio $\frac{b}{p_{jt}h}$. This specification captures the empirical regularity that households borrowing a larger share of their housing wealth face higher interest rates, consistent with evidence on loan-to-value pricing documented in Section 2.

In each period t between shock ages, liquid wealth evolves according to:

$$\dot{b}_t = y_t(j,s,a) + q_t(b,j,h)b_t - c_t - (r_{jt} + p_{jt}\tau^r_{h,at})h^r_t - (\delta + \tau^o_{h,at})p_{jt}h + \tau_{jat}.$$

The right-hand side captures labor income $y_t(j,s,a)$, net interest income (or cost) $q_t(b,j,h)b$ and liquid wealth b_t, consumption c_t, rental costs $(r_{jt} + p_{jt}\tau^r_{jat})h^r$, homeownership costs $(\delta + \tau^o_{jat})p_{jt}h$, and location-age-specific transfers τ_{jat}.

The parameters τ^r_{jat} and τ^o_{jat} represent proportional taxes on rented and owned housing, respectively, while τ_{jat} represents a location-age-specific lump-sum transfer (potentially negative). This specification departs from the standard spatial economics literature in three ways, each motivated by UK institutional features. First, we impose taxes on both renters and owners because all UK residents pay council tax regardless of tenure status. Second, we include a lump-sum component τ_{jat} to capture the regressive structure of council tax, which imposes a high floor even on low-value dwellings. Third, allowing location and age variation in τ_{jat} provides flexibility to evaluate place-age-based policies that target young workers in productive cities.

At shock ages, individuals make discrete adjustments to their liquid wealth and housing. The decision sequence proceeds as follows. First, immediately after choosing their new residence location j', individuals sell any existing owner-occupied housing in their previous location j (if relocating) and update their financial position:

$$\begin{aligned}\tilde{b}^R(b,h,j;j') &= b + \mathbf{1}(j \neq j')\left[(1-\Psi)p_{jt}h\right];\\ \tilde{h}^R(h,j;j') &= \mathbf{1}(j = j')h.\end{aligned}$$

If relocating ($j \neq j'$), the individual receives net proceeds $(1-\Psi)p_{jt}h$ from the sale and arrives at the new location without housing. If staying ($j = j'$), existing housing is retained.[23]

Second, individuals choose housing tenure (rent versus own) and size. For owners, this choice determines $h' \in \mathscr{H}^o$, while for renters $h' = 0$. After this housing decision, liquid wealth adjusts to:

$$\tilde{b}^H(b,h,j;h') = b + \mathbf{1}(h' \neq h)\left[(1-\Psi)p_{jt}h - p_{jt}h'\right],$$

[23]The individual state variables for b and h, therefore, update accordingly.

where any change in owner-occupied housing ($h' \neq h$) involves selling the existing property and purchasing the new one. The set of feasible housing choices is constrained by borrowing limits:

$$\mathscr{H}_t(b,h,s,a,j) = \left\{h' \in \mathscr{H}^o \cup \{0\} : \tilde{b}^H(b,h,j;h') \geq \max\{-M_p p_{jt} h', -M_y y_t(j,s,a)\}\right\}. \quad (4)$$

This constraint ensures that individuals can only adjust housing if the resulting wealth position satisfies both the loan-to-value and income multiplier constraints. Consequently, borrowing constraints are never violated at shock ages by construction. An important implication is that households in negative equity—where selling their property would leave them with negative liquid wealth—face severe mobility restrictions, as they cannot afford to relocate or adjust their housing without violating borrowing constraints.

4.2.6 Household Value Function

We denote the vector of individual state variables as $\Omega := (b,h,j,s,a)$, where b is liquid wealth, h is owner-occupied housing size, j is residence and work location, s is the idiosyncratic skill level, and a is age.

Since the bequest motive depends only on liquid wealth, households optimally liquidate all owner-occupied housing at death. The value function at the terminal age $\overline{A}$ is therefore evaluated at the post-sale wealth level:

$$V_t(b,h,j,s,\overline{A}) = v(b + (1-\Psi)p_{jt}h).$$

Between shock ages, individuals solve a standard consumption-saving problem conditional on their current location j and housing stock h. The value function satisfies the Hamilton-Jacobi-Bellman (HJB) equation:

$$\rho V_t(\Omega) = \max_{c,h^r} u(c, \mathbf{h}(h,h^r)) + \partial_b V_t(\Omega)\dot{b}(\Omega,c,h^r) + \partial_a V_t(\Omega) + \partial_t V_t(\Omega), \quad (5)$$
$$\text{s.t.} \begin{cases} \dot{b} \geq 0 \text{ if } (b \leq -M_p p_{jt} h) \vee (b \leq -M_y y_t(j,s,a)), \\ h^r = 0 \text{ if } h > 0. \end{cases}$$

The first constraint requires non-negative saving when borrowing limits bind, while the second constraint prohibits simultaneous renting and owning. The partial derivative $\partial_b V_t(\Omega)$ represents the marginal value of wealth, $\partial_a V_t(\Omega)$ captures the direct effect of aging on utility, and $\partial_t V_t(\Omega)$ reflects changes in the value function due to time-varying aggregate conditions (prices and wages).

At shock ages, the value function is determined through a backward recursion over the sequence of decisions. We solve this recursion in three stages, working backward from housing choice to location choice to skill realization.

Stage 3: Housing choice. After observing the skill shock and choosing location j, individuals select their optimal housing tenure and size. Denoting $\mathscr{H}_t(\Omega)$ the set of feasible housing choices defined in Equation (4), households draw an idiosyncratic preference shock over each of the feasible housing choices, denoted $\{\varepsilon_{h'}\}_{h' \in \mathscr{H}_t(\Omega)}$ with scale parameter ν_h.

Let $V^H(\Omega)$ denote the expected value of housing choice faced by the agent immediately before his choice, and after the realisation of the idiosyncratic shocks. Given the Gumbel distribution of preference shocks, the expected value at this stage takes the closed form:

$$V_t^H(\Omega) = \nu_h \ln \left\{ \sum_{h' \in \mathscr{H}_t(\Omega)} \exp \left[\frac{1}{\nu_h} V_t(\tilde{b}^H(\Omega; h'), h', j, s, a) \right] \right\}.$$

Stage 2: Location choice. Before choosing location but after observing the skill shock ζ, individuals draw idiosyncratic preference shocks $\{\varepsilon_{j'}\}_{j' \in J}$ and select their optimal residence location. Given the Gumbel distribution of preference shocks, the expected value at this stage takes the closed form:

$$V_t^R(\Omega) = \nu \ln \left\{ \sum_{j' \in J} \exp \left[\frac{1}{\nu} \left(V_t^H(\tilde{b}^R(\Omega; j'), \tilde{h}^R(\Omega; j'), j', s, a) + \mu_{j'}(a) - \kappa_{j j'} \right) \right] \right\}.$$

Stage 1: Skill realization. At the beginning of shock ages, before observing the productivity shock, the expected value depends on whether the individual is a worker or retiree:

$$V_t(\Omega) = \begin{cases} \int V_t^R(b, h, j, s', a)\, dF_j(s'|s, a) & \text{if } a < A_{\text{ret}}, \\ V_t^R(\Omega) & \text{otherwise}, \end{cases} \tag{6}$$

where $F_j(s'|s,a)$ is the distribution of next-period skill s' conditional on current skill s, age a, and location j (reflecting the location-specific learning process defined in equation (3)). For workers, the value integrates over possible skill realizations. For retirees, skills are fixed and the value reduces to the location choice problem.

4.2.7 Initial Conditions

Households enter the economy at age $a = 0$ with an initial joint distribution of liquid wealth, skill, and location (b, s, j) characterized by density $g_0(b, s, j)$. We allow this initial distribution to depend on the contemporaneous location and skill distribution of other age groups, capturing

potential intergenerational linkages such as parental wealth transfers or inherited human capital. All individuals begin life with zero owner-occupied housing ($h = 0$). The model treats age zero as a shock age, meaning that newborns immediately make location and housing decisions. They may relocate from their initial location by paying the age-specific moving cost and they choose their initial housing tenure and size from the feasible set.

4.2.8 Density of State Variables

Between shock ages, the density of state variables $g_t(\Omega)$ evolves according to the Kolmogorov Forward Equation (KFE):

$$\partial_t g_t(\Omega) = -\partial_b \left[\dot{b}(\Omega) g_t(\Omega)\right] - \partial_a g_t(\Omega). \tag{7}$$

The first term captures the redistribution of mass across wealth levels due to savings flows $\dot{b}(\Omega)$, while the second term captures deterministic aging.

At shock ages, the distribution undergoes discrete changes as individuals experience skill shocks, relocate, and adjust housing. We track these changes sequentially through each decision stage. Let x^- denote the value of state variable x immediately before the shock age.

Stage 1: After skill realization. Immediately after productivity shocks are realized but before location and housing choices, the density becomes:

$$g_t^s(\Omega) = \begin{cases} \sum_{j^-} \int \lim_{\iota \to 0} g_{t-\iota}(b,h,j^-,s^-,a-\iota)\, f_{j^-}(s|s^-,a)\, ds^- & \text{if } a < A_{\text{ret}}, \\ \lim_{\iota \to 0} g_{t-\iota}(b,h,j,s,a-\iota) & \text{otherwise.} \end{cases}$$

For working-age individuals, the density integrates over all possible previous skill levels s^-, weighting by the location-specific transition probability $f_{j^-}(s|s^-,a)$. The sum over j^- accounts for individuals currently in location j who worked in different locations j^- during the previous period. For retirees, skills remain fixed.

Stage 2: After location choice. After individuals observe location preference shocks and choose their residence location, the density becomes:

$$g_t^R(\Omega) = \sum_{j^-} \iint \mathbf{1}_t^R(b^-,h^-,j^-;b,h,j)\, \pi_t^R(b^-,h^-,j^-,s,a;j)\, g_t^s(b^-,h^-,j^-,s,a)\, db^- dh^-,$$

where $\mathbf{1}_t^R(b^-,h^-,j^-;b,h,j)$ is an indicator function equal to one when the post-relocation state (b,h,j) corresponds to the pre-relocation state (b^-,h^-,j^-) after applying the relocation

rules:

$$\mathbf{1}_t^R(b^-,h^-,j^-;b,h,j) = \mathbf{1}\left(b = \tilde{b}_t^R(b^-,h^-,j^-;j) \wedge h = \tilde{h}^R(h^-,j^-;j)\right).$$

The relocation probability $\pi_t^R(\Omega;j')$ represents the fraction of households with pre-relocation state $\Omega = (b^-,h^-,j^-,s,a)$ who choose destination j'. Given the Gumbel distribution of preference shocks, this probability takes the logit form:

$$\pi_t^R(\Omega;j') = \frac{\exp\left[\frac{1}{\nu}\left(V_t^H(\tilde{b}^R(\Omega;j'),\tilde{h}^R(\Omega;j'),j',s,a) - \kappa_{jj'}\right)\right]}{\sum_{j''}\exp\left[\frac{1}{\nu}\left(V_t^H(\tilde{b}^R(\Omega;j''),\tilde{h}^R(\Omega;j''),j'',s,a) - \kappa_{jj''}\right)\right]}. \tag{8}$$

Stage 3: After housing choice. Finally, after choosing housing tenure and size h, which may involve adjusting wealth b, the post-shock-age density is:

$$g_t(\Omega) = \iint \mathbf{1}_t^H(b^-,h^-,j,s,a;b,h)\, g_t^R(b^-,h^-,j,s,a)\, db^- dh^-,$$

where $\mathbf{1}_t^H(b^-,h^-,j,s,a;b,h)$ is an indicator equal to one when the post-housing-choice state (b,h) corresponds to the optimal choice given pre-choice state (b^-,h^-):

$$\mathbf{1}_t^H(b^-,h^-,j,s,a;b,h) = \mathbf{1}\left(b = \tilde{b}^H(b^-,h^-,j;\tilde{h}(b^-,h^-,j,s,a)) \wedge h = \tilde{h}(b^-,h^-,j,s,a)\right). \tag{9}$$

Aggregation. Given the joint distribution $g_t(\Omega)$ of state variables, we can compute aggregate quantities in each location. Labor supply in location j equals the total efficiency units supplied by working-age residents:

$$L_{jt} = \int_{a\in[0,A_{\text{ret}})} \left(\iiint e^s g_t(b,h,j,s,a)\, db\, dh\, ds\right) da. \tag{10}$$

Housing demand aggregates rental and owner-occupied housing:

$$H_{D,jt} = \iiiint \left[h_t^r(b,h,j,s,a) + h\right] g_t(b,h,j,s,a)\, db\, dh\, ds\, da. \tag{11}$$

Population in location j is:

$$N_{jt} = \iiiint g_t(b,h,j,s,a)\, db\, dh\, ds\, da.$$

4.3 Production

The economy features two production sectors: a traded-goods sector and a housing sector.

4.3.1 Traded-Good Firms

The traded-goods sector produces a homogeneous output that can be freely shipped between locations and serves as the numeraire. Production uses only labor as an input. We assume perfect competition, so wages equal the marginal product of labor. The production function in location j at time t is:

$$Y_{jt} = Z_j L_{jt}^{\alpha}, \tag{12}$$

where L_{jt} is labor input measured in efficiency units, $\alpha \in (0,1)$ governs returns to scale, and Z_j is the innate total factor productivity. Profit maximization by competitive firms implies wages equal the marginal product of labor:

$$w_{jt} = \alpha Z_j L_{jt}^{\alpha-1}.$$

Inverting this wage equation yields labor demand:

$$L_{jt} = \left(\frac{\alpha Z_j}{w_{jt}}\right)^{\frac{1}{1-\alpha}}. \tag{13}$$

4.3.2 Housing

Developers construct new residential floorspace according to the production function:

$$Y_{jt}^h = \mathscr{F}(Z_j^h, K_{jt})$$

where Z_{jt}^h is a sector and location-specific productivity parameter, $K_{j,t}$ is capital input, measured in terms of units of traded good, $\mathscr{F}(\cdot)$ is a generic production function. Because we assume that the economy is small and open and there is perfect capital mobility, then the return of debt-free risk-free bond, q, is assumed to be exogenous and equal to rental rate of capital. Free capital mobility implies that any demanded level of capital will be satisfied, if it exceeds asset demanded by the households. In this context, we assume developers are profit-maximizing and respond to housing prices, so that the optimal construction level Y_{jt}^h is a non-decreasing function of the housing price p_{jt}. The specific functional form and microfoundation of the housing supply function are detailed in Appendix **??**, but the key property of the function $\mathscr{F}(\cdot)$ is that higher prices induce weakly more construction: $\frac{\partial Y_{jt}^h}{\partial p_{jt}} \geq 0$.

Housing construction is irreversible, so $Y_{jt}^h \geq 0$. The total stock of residential floorspace in

location j evolves according to:

$$\dot{H}_{jt} = Y^h_{jt} - \delta H_{jt}, \tag{14}$$

where δ is the depreciation rate. New construction Y^h_{jt} adds to the housing stock, while depreciation erodes it at rate δ.

4.3.3 Real Estate Agents

Rental housing is supplied by perfectly competitive real estate agents who act as intermediaries between developers and renters. There is free entry. These agents purchase properties at price p_{jt}, finance the purchase by borrowing from an external credit market at interest rate q, and rent the properties to households at rental rate r_{jt}. Operating rental properties entails management costs equal to a fraction m of the property value, capturing expenses for marketing, tenant placement, maintenance coordination, and property management.[24] Perfect competition drives expected profits to zero. The zero-profit condition requires that rental income covers the opportunity cost of capital qp_{jt}, depreciation, and management costs. Additionally, agents must account for expected capital gains (or losses) from housing price appreciation $\dot{p}_{jt}$, which reduces (or increases) the required rental rate. The equilibrium rental rate therefore satisfies:

$$r_{jt} = (\delta + q + m)p_{jt} - \dot{p}_{jt}.$$

The management cost parameter m creates a wedge between the cost of owning and renting, contributing to the age-specific housing cost differential documented in Section 2.

4.4 Equilibrium

An *Equilibrium* is an allocation, household value function $V_t(\Omega)$, density of state variables $g_t(\Omega)$, wages w_{jt}, rents r_{jt} and house prices p_{jt} such that:

1. Households optimise, satisfying Equations (5) to (6);
2. The density of state variables is consistent with households' optimisation: $g_t(\Omega)$ satisfies Equations (7) to (9);
3. Firms maximise profits, consistently with (12) to (13)
4. Labour market clears: labour demand (13) equals labour supply (10) in all locations
5. The housing market clears: housing stock (14) equals demand (11) in all locations

[24] In practice, these costs reflect both direct management expenses and the time and effort required to maintain rental properties, including dealing with tenant turnover, repairs, and regulatory compliance.

A *Stationary Equilibrium* is an Equilibrium where all equilibrium objects are time-invariant.

4.4.1 Consumption-Saving Problem Characterisation

Since our utility function is a Cobb-Douglas for the intra-temporal consumption/housing problem nested inside a CRRA for the inter-temporal expenditure-saving problem, we divide the consumption-saving problem into two separate parts.

Intra-temporal problem. At each moment in time, for a given expenditure level e_t, the individual decides to allocate their consumption to either the consumption good c_t or housing.

If the individual owns a house ($h_t > 0$), all expenditure goes to consumption since the level of housing is fixed for the period. Hence, the solution to the consumption/housing problem is simply $c_t = e_t$ and $\approx_t = \chi h_t$.

If the individual is renting, then the intra-temporal housing problem solves a simple Cobb-Douglas consumption problem with total expenditure e_t. Hence, the individual spends σe_t on the consumption good and $(1-\sigma)e_t$ on the housing good. Hence, $c_t = \sigma e_t$ and $\approx_t = (1-\sigma)e_t r_{jt}^{-1}$.

Thus, we can write the indirect utility function for houseowners as

$$\hat{U}_t(e_t, h_t, j) = \begin{cases} \mu_j \dfrac{\left(e_t^{\sigma}(\chi h_t)^{(1-\sigma)}\right)^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}} & \text{if } h_t > 0, \\ \mu_j e_t^{1-\frac{1}{\theta}} \dfrac{\left(r_{jt}^{\sigma-1}\right)^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}} & \text{if } h_t = 0. \end{cases}$$

Inter-temporal problem. Individuals optimally decide how much to spend and save across periods. Away from the constrained points, the problem is characterised by the first-order condition of their inter-temporal saving problem:

$$\hat{U}'_e(e,h,j) - \partial_b V(b,h,j,a) = 0,$$

which is characterised as

$$\begin{cases} \mu_j e_t^{-\frac{\sigma}{\theta}} (\chi h)^{-\frac{(1-\sigma)}{\theta}} = \partial_b V(b,h,j,a) & \text{if } h_t > 0, \\ \mu_j e_t^{-\frac{1}{\theta}} \left(r_{jt}^{\sigma-1}\right)^{1-\frac{1}{\theta}} = \partial_b V(b,0,j,a) & \text{if } h_t = 0. \end{cases}$$

Finally, at the borrowing-constrained points the solutions to the consumption-saving problems

satisfy

$$\partial_b V(b,0,j,a) \geq \partial_c U(c,h^r)\partial_b c^*(\cdot) + \partial_{h^r} U(c,h^r)\partial_b h^{r,*}(\cdot), \forall b \leq 0, \text{if } h = 0,$$
$$\partial_b V(b,h,j,a) \geq \partial_c U(c,\mathbf{h}), \text{if } h > 0,$$

with $\partial_b c^*(\cdot) + \partial_b h^{r,*}(\cdot) r_j = 1$.[25]

4.4.2 Endogenizing Age Differences in Housing Costs

The tractable model in Section 3 assumed an exogenous housing cost differential ψ between young and old workers. The quantitative model endogenizes this differential through two complementary mechanisms that interact with life-cycle patterns in wealth accumulation.

Direct mechanism: Leverage-dependent mortgage rates. The interest rate function $q_t(b,j,h) = q + \mathbf{1}[b<0] f(b, p_{jt}h)$ creates age-specific financing costs. Young workers, who typically have low wealth, must borrow a larger fraction of the property value (high leverage $|b|/(p_{jt}h)$) and consequently face higher interest rates through the increasing function $f(\cdot)$. This generates higher per-period mortgage costs for young homeowners. In contrast, older workers who have accumulated wealth can make larger down payments, achieve lower leverage ratios, and secure more favorable interest rates. Additionally, older workers who purchased homes in the past have built up equity through mortgage repayment and house price appreciation, further reducing their borrowing costs when adjusting housing.

Indirect Mechanism: Tenure choice and the rental premium. Young and old workers face different trade-offs when choosing between renting and owning, leading to age-specific tenure patterns that affect average housing costs. Young workers are more likely to rent for three reasons: (i) they face binding borrowing constraints due to low wealth and income, (ii) they pay higher mortgage rates due to high leverage (Direct Mechanism), and (iii) they anticipate higher mobility, both across locations and housing sizes, making the transaction costs of home sales, Ψ, more burdensome in expectation. However, rental rates incorporate the management cost m, creating a rental premium. Our calibration in Section 5 confirms that $m \approx q_2(M_p)$, where $q_2(M_p)$ is the interest rate premium at maximum leverage M_p. This implies that renting is as expensive as the highest-leveraged mortgage. This cost difference persists even after accounting for the opportunity cost of the minimum down payment $(1-M_p)p_{jt}$ and the amortized transaction costs from less frequent home sales.

These two mechanisms interact with life-cycle wealth accumulation to generate the age-varying housing costs documented in Section 2. Young workers face high housing costs because they

[25] This derivative simply states that any change in savings is fully distributed across higher or lower consumption and housing services purchase, when the individuals are renting.

must either rent (paying the management premium m) or purchase with high leverage (paying elevated interest rates). Older workers face lower costs because they can purchase with low leverage (low interest rates) or have already repaid a large share of their mortgage, reducing the interest rate they pay on the remaining part. The model thus microfounds the reduced-form parameter ψ from the tractable model while preserving its key implication: systematic age differences in the price of urban space create inefficient spatial sorting.

5 Estimations and Results

Locations We define a location in the model as a commuting zone (TTWAs in the data). Using the total number of TTWAs is computationally challenging. Moreover, not all of the original 173 Travel-to-Work Areas (TTWAs) have sufficient sample sizes, and some, especially when small, exhibit high rates of cross-boundary commuting. We apply an algorithm to merge smaller TTWAs with neighbouring areas, preserving economic and demographic features. This results in 52 spatial units of sufficiently large size and internal commuting cohesion.[26]

In Table 4 we list the externally calibrated parameters used to estimate the model. We estimate all parameters from UK data sources, besides three. First, we set Ψ, the transaction cost of selling a house, to 0.06, as standard in the literature (Greaney et al., 2025). Second, we set θ_1, the curvature parameter of the warm glow bequest function, to 0.01 as in (Greaney et al., 2025) to approximate a log-linear bequest motif while preserving a well-characterised terminal condition. Third, since no clear data is available for this parameter, we set $M_y = 4.0$. This takes into account how UK lenders quote as maximal lending multiplier values around 4.0-4.5, with financial regulations suggesting that - on top of this multiple - households should also be able to withstand interest rates stress tests. We provide further Details in Appendix C.

In Table 5 we list the parameters we estimate via Simulated Method of Moments. For each parameter, we list its target, the value of the target in the data, the one simulated in the model, and - where possible - the estimated value of the parameter. Overall, we match 7 "general" moments, 51 amenities (the number of location, minus one normalised to 1), and 52 parameters

[26]The algorithm groups together bordering TTWA with less than 350,000 workers, starting from the smallest TTWA (with the expection of the functionally isolated Isle of Wight). We iteratively merge any TTWA below the threshold with one neighbouring TTWA (neighbourhood defined by shared polygon borders using the ONS TTWA 2011 shapefile). When choosing the merge partner, we restrict attention to adjacent TTWAs that are similar in Value added per worker and return to experience (measured as average wage growth of stayers). In particular, we only allow the merge of locations are are either: 1) both above the mean in VA and experience; 2) both below the mean; 3) both above the mean in VA and below the mean in experience; 4) both below the mean in VA and above the mean in experience. If available, we further restrict our attention to the set of neighbours that exhibit strong commuting ties, measured by the share of commuters from the origin TTWA to the candidate destination exceeding 4% of the origin's resident population. Among the remaining candidates, the merge partner is the smallest eligible neighbour (by population), and the two areas are collapsed into a single macro TTWA. The algorithm is repeated until all new units satisfy the dimension treshold. The new spatial units have on average 76% of internal commuting.

for labour and housing productivity, for a total of 162 parameters. While the number of parameters is particularly large, de facto we use 156 of them (amenities, homeownership preference and productivities) to perfectly match the observed population shares, average homeownership, wages and prices given all other parameters.

Table 4: Externally Calibrated Parameters

Parameter	Description	Target or Source	Value	Reference
Externally Calibrated				
M_p	Collateral constraint	BoE maximum LTV	0.90	Appendix C.1
M_y	Income multiplier constraint	UK standard	4.0	Appendix C.1
q	Risk-free rate	Data	2.07%	Appendix C.1
q_1	Loan rate premium	BoE mortgage rates	1.343%	Appendix C.1
q_2	Unsecured loan premium	BoE 60% vs 90% LTV premium	5.240%	Appendix C.1
δ	Housing depreciation	ONS National Accounts	0.895%	Appendix C.2
m	Estate Agents costs	Ownership-rent cost gap	0.0158	Appendix C.3
Ψ	Housing transaction fee	Standard value	0.06	Greaney et al. (2025)
$H_{\min}$	Minimum house size	1st percentile house sizes	1.97	Appendix C.5
$H_{\max}$	Maximum house size	90th percentile house sizes	3.75	Appendix C.5
θ	CRRA curvature	Intertemporal Elasticity of Substitution	0.5	Standard value
σ	Share of consumption	Expenditure on non-housing	0.72	Appendix C.5
α	Returns to labour input	Labour share of production	0.61	ONS
D	Distance matrix	Travel times		Appendix C.6
θ_1	Bequest motive curvature	Log-linear motive	0.01	Greaney et al. (2025)
θ_s	Labour prod. persistence	ASHE		Appendix C.7
σ_s	Labour prod. variance	ASHE		Appendix C.7
γ	Wage growth parameters	ASHE		Appendix C.7
$g_0(b,s,j)$	Initial wealth, skill, location	Varies		Appendix C.8
τ_y	Payroll taxes	UK average payroll tax	0.294	OECD
τ^r, τ^o	Council tax - curvature	Share of value at max H	0.002	Appendix C.4
τ	Council tax - base	Minimum council tax	0.025	Appendix C.4
τ_a	Age part of council tax	Discount by age	0.0	Appendix C.4

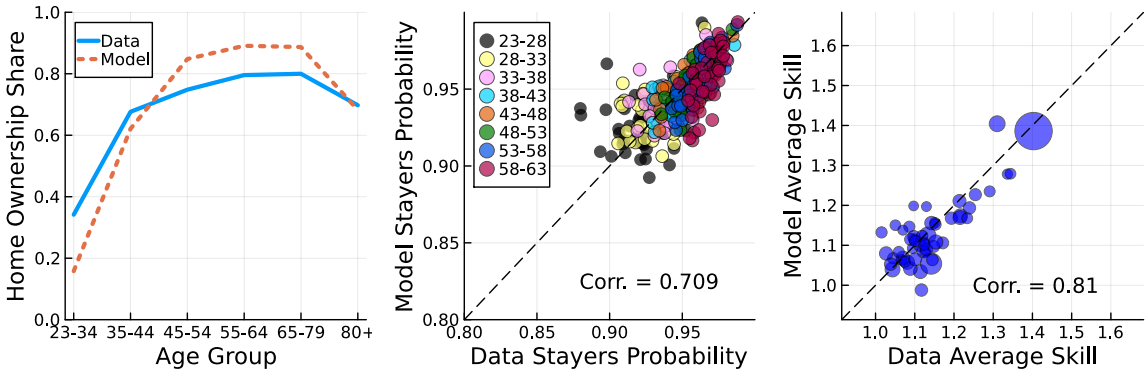
Table 5: Parameters Calibrated with Indirect Inference

			Moment Value			
Parameter	Description	Target	Data	Model	Parameter Value	Reference
ρ	Discount rate	Income/wealth ratio	3.15	3.15	0.034	Appendix C.9
ν	Location taste shock scale	Relocation flows	varies		3.200	Appendix C.10
μ_j	Amenities	Residence shares	varies		varies	Appendix C.10
κ	Moving cost, intercept	Relocation flows	varies		5.303	Appendix C.10
κ^d	Moving cost, distance	Moving - distance gradient	-1.13	-1.13	3.326	Appendix C.10
κ^a	Moving cost, age	Moving - age gradient	-0.014	-0.014	0.0095	Appendix C.10
χ	Homeownership preference	Homeownership share	0.674	0.674	1.111	Appendix C.11
θ_0	Bequest motive	Homeownership share, 80+ y.o.	0.697	0.697	9.001	Appendix C.12
Z_j	Innate productivity	Wages	varies		varies	Appendix C.13
Z_j^h	Housing productivity	House prices	varies		varies	Appendix C.14

5.1 Non-Targeted Moments

In this section, we evaluate the ability of our model to match moments we did not target in our estimation.

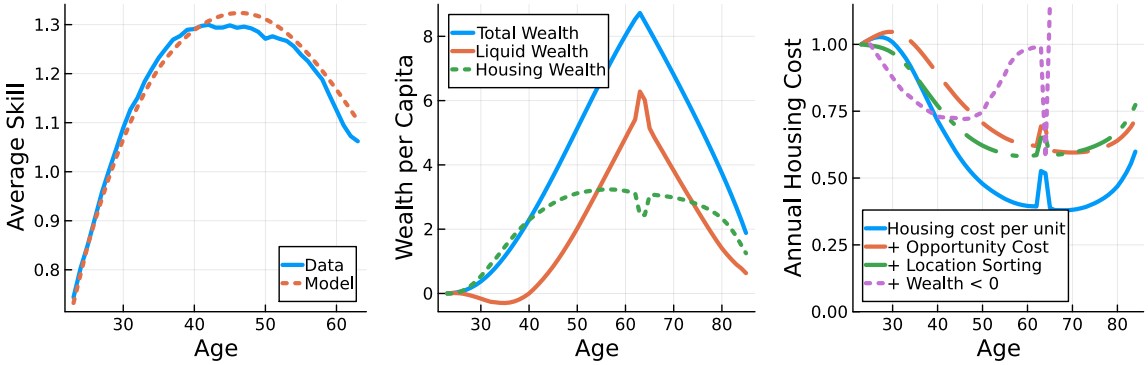
Figure 7: Non-Targeted moments and model fit



(a) Homeonwership, by age

(b) Stayers, by age

(c) Avg. Skill, by location



(d) Avg. skill, by age

(e) Wealth, by age

(f) Housing Cost Index, by age

Housing In panel (a), we plot the homeownership observed in the data (blue solid line) against the one simulated from the model (dotted orange line). Our model matches ownership at 35-44 years old and at 80+ years old (the latter being targeted), while undershooting the ownership of the young (25-34) and overshooting the one of the middle-aged (45-79). The former is likely due to not modelling that some workers have been working since the age of 18, so already have 5 years of seniority and savings by when they turn 23, and lack of inheritance; the latter is likely due to not having explicitly modelled extreme life income events that may prevent some workers from ever owning a house, which is always strictly preferable to renting as long as the individuals does not expect to be highly mobile, such as unemployment shocks.

Stayers, by age The model replicates well the share of stayers by age (panel b) with a correlation of 0.71, despite the parameters are estimate targeting only the overall stayers and the *dispersion* of movers by age (rather than their level). The errors are slightly larger for the young, but nearly all the simulated data points are within a reasonable +/- 5% distance from the data.

Skill-Location Sorting The model performs extremely well in matching the degree of sorting across locations (panel c). The distributions of average skills by location of ASHE data and our model have a correlation of 0.81. Moreover, the model reproduces well the *level* of skills, with the estimates sitting along the 45 degrees line despite not explicitly targeting this moment. In panel (d) we plot the average skill by age of working-age individuals. Again, the model matches the data well - in particular among the young - despite not having explicitly targeted the average level or sorting of skills.

Wealth-age Profile In panel (e) we show the evolution of wealth over the age profile. Individuals rapidly accumulate housing in their youth, while holding negative liquid wealth balances (due to mortgages). While our model does not match the depth of negative wealth balances around the age of 40, recall that our model does not model retirement private savings. That is, we do not model that part of the wealth available to individuals may be held in accounts that cannot be used towards one's mortgage.

Housing Cost-Age Gradient Finally and most importantly, we show that the model replicates well the magnitude of the housing cost-age gradient observed in the data. Our model predict that the average middle-aged, aged 45 to 64, pays on average 0.361 log-points less for an equivalent housing unit, relative to a young worker aged 25-34. When we only look at homeowners, we find a difference in housing costs of 0.22 log-points, and when we restrict our sample to homeowners with negative wealth, we find a difference of 0.15 log points in favour of the old. Our model matches well both the overall housing cost gap (blue solid line in panel f) and the housing gap among mortgagers (dotted purple line).

6 Counterfactuals and Policies

6.1 Place-Age-Based Policies

In this section, we evaluate a policy that directly targets the housing-cost gap between young and old workers in productive locations. Specifically, we simulate a place-age-based tax schedule that reduces property and service taxes for young individuals aged 21–34 in the five labour markets with above-average productivity, financed by a revenue-neutral increase for individuals aged 40+. In the UK, property and service taxes are unified under the "council tax" and paid by the occupier rather than the owner; accordingly, we apply the same age-based discount to renters as to owner-occupiers.[27]

[27] We define revenue-neutrality as the sum of the changes in all taxes collected across all locations — including both payroll and property taxes — summing to zero.

The direct effect of this policy is to compress the housing-cost gap between young and old workers in productive labour markets relative to less-productive ones. We find that a council tax discount for the young equivalent to 0.6% of the house price — equivalently, 4.1% of the average rent, or an 88% discount on the existing council tax for the lowest-valued properties — combined with an increase of 0.12% of the property value for the old, would fully equalise the housing-cost age gap across productive and unproductive locations. By differentially reducing the relative cost of productive locations for the young while raising it for the old, this schedule strengthens the incentive to sort into cities early in the life cycle and relocate out at older ages, thereby stimulating skill accumulation and raising aggregate output.

Figure 8 plots the results for different policy sizes. The policy that equalises city-rural housing-cost gaps increases output by nearly 0.1%. A policy resulting in approximately an 18% discount on the combined rent and council tax for the young (equivalent to 2% of the value of the house) would achieve a 0.23% output gain. Beyond output gains, this policy also delivers important welfare gains.[28] In consumption-equivalent terms, the gap-equalising policy increases welfare by nearly 0.3%, rising to 0.9% under a subsidy of 2% of the house value.

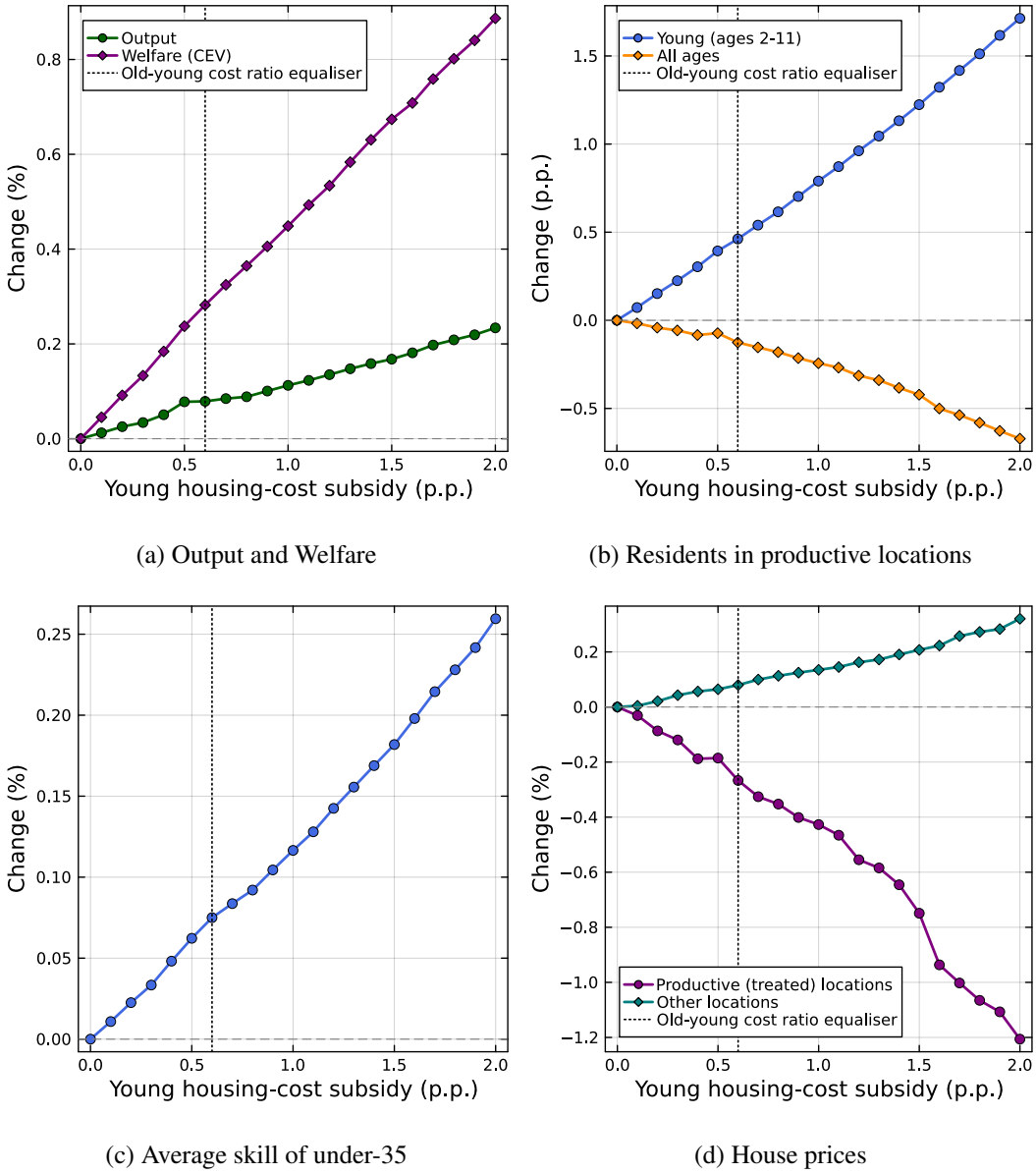
The core mechanism behind these results is illustrated in Figure 8b. While the equalising policy slightly reduces the share of individuals living in productive locations by 0.1 percentage points (from 21.9 to 21.8% of all individuals), the share of young workers residing in productive locations increases by 0.5 p.p. (from 29.8 to 30.3% of all young). This implies that the young accumulate more skills in equilibrium (Figure 8d), and are also more likely to be located in higher-wage locations, which become relatively cheaper (Figure 8c) as older individuals — who also demand more housing — move out faster than younger individuals do. Notice that prices in productive locations fall by about twice as much as the additional tax applied on older individuals, due to lower-skilled individuals moving into productive locations, which reduces the average skill accumulation of incumbents (but increases it in aggregate).

Finally, we show that these results do not arise from taxing unproductive retirees. To show this, we repeat the exercise of housing subsidies in productive locations for the young, but increase taxes in a revenue-neutral manner only for individuals aged 40–64. That is, we do not increase taxes for retirees living in productive locations. We show the results for output and welfare in Figure 9. We find that, depending on the policy size, between 75% and 90% of the output gains are preserved. Welfare gains are between 7% lower and 5% larger.[29] These results show that

[28] We calculate steady-state welfare gains as the percentage increase in the consumption bundle $c^{\sigma}\mathbf{h}^{1-\sigma}$ which, when applied to every period of life of an agent, would make an individual of age 0 indifferent between the old and new equilibrium. Such increase $x(b,h,j,s)$ differs for every level of wealth b, housing h, location j and skill s; we therefore calculate aggregate welfare gains as the average $x(b,h,j,s)$ across all individuals. A derivation is provided in Appendix E.1.

[29] The reason welfare gains can be larger despite a fall in output is that some retirees are wealthy hand-to-mouth consumers with high MPCs. Taxing individuals during the old age, as done in baseline counterfactual but not the

Figure 8: Policy effects of a place-age-based tax policy, by policy size



(a) Output and Welfare

(b) Residents in productive locations

(c) Average skill of under-35

(d) House prices

Notes: The figure plots the steady-state effects of different policies that reduce housing costs for young workers (aged 21-34) in productive labour markets, financed by a revenue-neutral increase in the equivalent tax for individuals aged 40+. On the x axis, we plot the size of the discount for the young as a percentage of house prices.

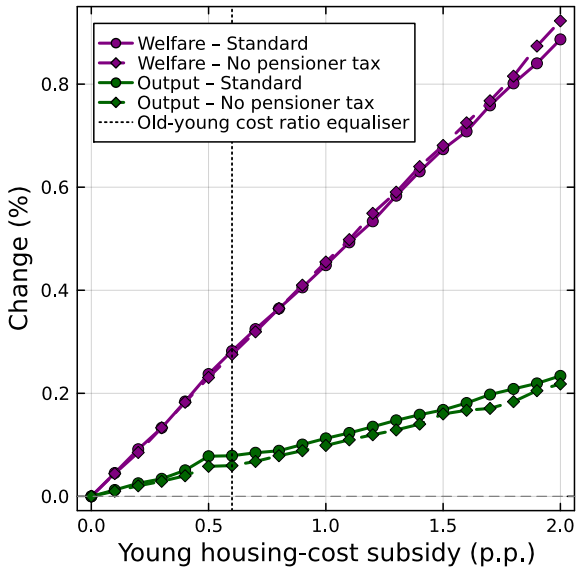
directly incentivising the relocation of retirees is *not* the main driver of our findings.[30] This is important because it shows that the mechanism highlighted in our model — competition for space among individuals with different *learning* rates — is distinct from competition between productive and non-productive individuals. In fact, in our quantitative model there are not only too many unproductive pensioners with low housing costs (due to previous purchases), but also

alternative discussed in this paragraph, can be very costly in welfare terms.

[30]Notice that taxing individuals before retirement can still lead to fewer retirees living in cities due to moving costs and housing transaction costs: individuals who move out of the city are unlikely to move back when older.

too many low-learning, high-productivity workers (the old).

Figure 9: Output and Welfare counterfactual gains, no tax increase for retirees



Notes: The figure plots the steady-state effects of different policies that reduce housing costs for young workers (aged 21-34) in productive labour markets, financed by a revenue-neutral increase in the equivalent tax for individuals aged 40-64. The results are compared with the gains from an equivalent counterfactual where the tax on older individuals is applied also to 65+ years old.

7 Conclusion

Cities provide both higher wages and higher returns to experience. In this paper, we have shown how this dual labour market role of cities (as places where to better employ skills, and where to accumulate skills faster) is important for both present and future aggregate productivity. However, when not all agents care about current and future wages in the same way, the composition of cities can be inefficient.

Since old workers do not care about returns to experience, their incentives are not aligned with those of younger generations. However, both young and old compete for the same limited city space. This generates what we call an *Spatial Inter-generational Congestion*.

Using a spatial OLG model, we show that - under plausible conditions on the marginal cost of housing for different generations and the spatial distribution of production and experience - the presence of spatial inter-generational congestion leads to under-accumulating human capital in the market equilibrium. A planner would allocate more young to cities than the decentralised equilibrium. This misallocation is increasing in the spatial correlation and size of spatial productivity and experience gaps. Unlike previous contributions to the literature, this misallocation is *dynamic*, does not rely on skill-sorting mechanisms, spatial frictions to relocation (besides commuting costs), or financial frictions.

Our results provide two important results for policymakers. First, we show that there exists a welfare-improving policy space of Place-*Age*-based policies, aimed at increasing the accumulation of human capital. On the contrary, nearly-"lump-sum" (poll) taxes for living in a labour market, such as the current Council Tax in the UK, as well as value-based but age-independent housing taxes, can exacerbate dynamic misallocation effects due to differences in age incentives and skills. Taxes not applied, or heavily discounted, on primary residences have a similar effect due to young workers being more likely to rent and, hence, be exposed to the tax pass-through.

Second, policies that increase access to productive labour markets by reducing the ratio between rent (or house prices) and wages in cities are effective at reducing the misallocation arising from inter-generational congestion only insofar they disproportionatelly favour the young. A simple policy such as exempting young workers from council taxes in the 7 most productive UK labour markets, financed by increasing such tax on over-40 individuals in the same locations, would improve output by nearly 0.1% and welfare by 0.22%. In the UK, providing young workers with a revenue-neutral housing subsidiy to reside in cities equivalent to 18% of their current housing costs can increase output by over 0.2% and welfare by nearly 1% in consumption-equivalent terms.

References

ACABBI, E. M., A. ALATI, AND L. MAZZONE (2024): "Human capital ladders, cyclical sorting, and hysteresis," *Bank of England, Staff Working Paper No. 1,077*.

ADAMS, W., L. EINAV, AND J. LEVIN (2009): "Liquidity constraints and imperfect information in subprime lending," *American Economic Review*, 99, 49–84.

ADMATI, A. AND M. HELLWIG (2014): "The Bankers' New Clothes: What's Wrong with Banking and What to Do about It: with a new preface by the authors," *Economics Books*.

ATTANASIO, O., R. BOTTAZZI, H. LOW, L. NESHEIM, AND M. WAKEFIELD (2012): "Modelling the demand for housing over the life cycle," *Review of Economic Dynamics*, 15, 1–18.

BENETTON, M., P. BRACKE, AND N. GARBARINO (2018): "Down payment and mortgage rates: evidence from equity loans," .

BESTER, H. (1985): "Screening vs. rationing in credit markets with imperfect information," *The American Economic Review*, 75, 850–855.

CHAMPION, T. (2016): "Population change and migration in Britain since 1991," in *New patterns of migration in Britain*, ed. by T. Champion and J. Falkingham, Routledge.

DE LA ROCA, J. AND D. PUGA (2017): "Learning by working in big cities," *Review of Economic Studies*, 84, 106–142.

DELL'ARICCIA, G., D. IGAN, AND L. LAEVEN (2012): "Credit booms and lending standards: Evidence from the subprime mortgage market," *Journal of Money, Credit and Banking*, 44, 367–384, doi:10.1111/j.1538-4616.2011.00491.x.

DIAMOND, P. (1965a): "National debt in a neoclassical growth model," *American Economic Review*, 55, 1126–1150.

DIAMOND, P. A. (1965b): "National debt in a neoclassical growth model," *American Economic Review*, 55, 1126–1150.

DIAMOND, R. (2016): "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 106, 479–524, doi:10.1257/aer.20131706.

FAJGELBAUM, P. D. AND C. GAUBERT (2020): "Optimal Spatial Policies, Geography, and Sorting." *Quarterly Journal of Economics*, 135.

GANONG, P. AND D. SHOAG (2017): "Why has regional income convergence in the US declined?" *Journal of Urban Economics*, 102, 76–90.

GLAESER, E. L. AND D. C. MARÉ (2001): "Cities and skills," *Journal of Labor Economics*, 19, 316–342.

GREANEY, B., A. PARKHOMENKO, AND S. . I. E. J. VAN NIEUWERBURGH (2025): "Dynamic Urban Economics," .

GUAITOLI, G. (2025): "Modelling the Aggregate Effects of Housing Supply Policies," Working Papers wpdea2516, Department of Applied Economics at Universitat Autonoma of Barcelona.

HAURIN, D. R., P. H. HENDERSHOTT, AND S. M. WACHTER (1996): "Borrowing constraints and the tenure choice of young households," *Journal of Housing Research*, 7, 137–154.

HSIEH, C.-T. AND E. MORETTI (2019): "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 11, 1–39.

HUGGETT, M. AND G. VENTURA (1999): "On the distributional effects of social security reform," *Review of Economic Dynamics*, 2, 498–531.

KAPLAN, G., K. MITMAN, AND G. L. VIOLANTE (2020): "The Housing Boom and Bust: Model Meets Evidence," *Journal of Political Economy*, 128, 3285–3345, doi:10.1086/708816.

KAPLAN, G. AND G. L. VIOLANTE (2014): "A model of the consumption response to fiscal stimulus payments," *Econometrica*, 82, 1199–1239.

KAPLAN, G., G. L. VIOLANTE, AND J. WEIDNER (2014): "The wealthy hand-to-mouth," *Brookings Papers on Economic Activity*, 2014, 77–138.

LA ROCA, J. D. AND D. PUGA (2017): "Learning by working in big cities," *Review of Economic Studies*, 84, 106–142.

LEE, N. AND P. SISSONS (2016): "The geography of wage inequality in the UK," *Regional Studies*, 50, 629–643.

LHUILLIER, H. (2024): "Should I stay or should I grow?" *Working Paper*.

LIU, L. (2022): "The demand for long-term mortgage contracts and the role of collateral," *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

MORETTI, E. (2011): "Local Labor Markets," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4 of *Handbook of Labor Economics*, chap. 14, 1237–1313.

——— (2012): *The new geography of jobs*, Houghton Mifflin Harcourt.

ORTALO-MAGNÉ, F. AND S. RADY (2006): "Housing market dynamics: On the contribution of income shocks and credit constraints," *Review of Economic Studies*, 73, 459–485.

PIAZZESI, M. AND M. SCHNEIDER (2016): "Housing and macroeconomics," *Handbook of macroeconomics*, 2, 1547–1640.

QI, M., J. YANG, AND X. ZENG (2009): "Loss given default of high loan-to-value residential mortgages," *Journal of Banking & Finance*, 33, 788–799, doi:10.1016/j.jbankfin.2008.09.011.

ROBACK, J. (1982): "Wages, rents, and the quality of life," *Journal of political Economy*, 90, 1257–1278.

ROSEN, S. (1979): "Wage-based indexes of urban quality of life," *Current Issues in Urban Economics*, 74–104.

SAMUELSON, P. A. (1958): "An exact consumption-loan model of interest with or without the social contrivance of money," *Journal of political economy*, 66, 467–482.

STEIN, J. C. (1995): "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects*," *The Quarterly Journal of Economics*, 110, 379–406, doi:10.2307/2118444.

A Empirical Evidence

A.1 Housing Dataset

In Table 6 we report the summary statistics for the English Housing Survey, for the subsample 2009-2013 where we counduct our main analysis. All statistics are unweighted. Overall, we have 67,268 observations. Of these, 50% represent renters, 19% households with an outstanding "original" mortgage (connected to the purchase of the property, rather than equity extraction), and 32% owners outright (regardless of whether they inherited, bought cash, or extinguished the original mortgage). Notice that the survey over-samples renters. Using the survey weights, renters represent 24% of the population, outstanding mortgages 40%, and owners outright 36%.

For 33,710 of these observations we have information on their house purchase. Of these, 25,954 did not buy their house outright, and for 22,033 we have complete information on their mortgages (deposit, duration) and house value at purchase and at the survey date. Many of these mortgagers repaid their loans by the time of the Survey. Hence, we restrict our sample to individuals under 65 (due to having very few observations above 65) who are currently repaying their mortgage, and for whom we have full information on the key variables connected to their purchase. This yields a sample of 8,107 mortgagers on which we perform our main analysis, and 8.087 for which we have information on their initial deposit paid. Of these, 109 pay both rent and mortgage due to having bought only a partial share of the house (e.g. "affordable housing" units, common in places such as London, where an Housing Association allows buyers to take up smaller mortgages by selling only part of the property, and charging below-market rent for the remaining share).

Table 6: Summary Statistics for English Housing Survey (2009-2013)

Statistic	N	Mean	St. Dev.	Min	Max
Number of adults	67,268	1.78	0.77	1	11
Number of household heads	67,268	1.54	0.50	1	2
Household joint income (£)	67,268	508.05	403.93	68.83	11,329.18
Tenure length (years)	67,268	11.40	9.24	0.00	25.00
Number of rooms	67,268	3.66	1.00	2	10
Number of bedrooms	67,268	2.66	1.00	1	9
Floor area (sq ft)	67,268	86.56	46.77	11.91	974.86
Year of purchase	33,710	1,994.39	13.09	1,936	2,012
Initial house value (£)	31,716	119,147.20	127,166.80	29	2,425,000
Current house value (£)	32,136	241,914.50	174,208.80	265	4,000,000
Housing cost (£ per week)	50,422	115.89	85.50	0.00	1,846.15
Mortgage duration (years)	22,033	16.31	11.07	0	50
Deposit to house value ratio	25,954	0.45	0.36	0.0001	1.00
Renter (1=yes)	67,246	0.50	0.50	0	1
Mortgager (1=yes)	67,246	0.19	0.39	0	1
Owner (1=yes)	67,246	0.32	0.46	0	1

A.2 Housing Cost Regressions: Robustness

A.2.1 LTV Ratios: Recent Mortgages

In Table 7 we show the same statistics on LTV ratios as Table 1, but for individuals who bought a house within the last 10 years. Focusing on recent purchases allows us to focus on house purchasing choices happening within or close to the current age groups of the individuals. For example, the results shown in the main text will include some now-old individuals who purchased their house when young, wrongly attributing their LTV to the behaviour at older ages. We find that nearly two-thirds of 45-64 y.o. individuals who recently concluded a purchase did so with cash (26.2%) or with a LTV ratio of less than 60% (33.9%).

Table 7: Share of homeonwers in each LTV category at mortgage origination, purchase within the previous 10 years

	LTV Ratio					
Age Group	Cash	$\leq 60\%$	60-70%	70-80%	80-90%	$\geq 90\%$
25-34	0.025	0.090	0.078	0.181	0.371	0.255
35-44	0.044	0.234	0.132	0.179	0.255	0.156
45-64	0.262	0.339	0.089	0.105	0.137	0.068
65+	0.884	0.082	0.005	0.009	0.012	0.008

Notes: The table shows the share of homeowners who bought a house at most 10 years before the survey, in each LTV category at the time of mortgage origination, by age group. Based on survey data from 2009 to 2013. Households are weighted by the numer of household heads. The shares are calculated based on the weights provided in the dataset. The LTV categories are defined based on the ratio of the deposit to the total value of the property financed by the mortgage, at the moment of purchase.

We then repeat the regression analysis of the housing cost gaps, using only information from recent ($\leq$10 years) mortgages. We report the results in Table 8. The estimate housing cost gap is slightly smaller (0.02 log-points) than in the main text, but the explanatory power of gaps in downpayments is slightly larger, with downpayments explaining up to 45% (0.05 log-points) of the difference in house prices between young and old. In Table 9 we further restrict the analysis to 5-years recent mortgages. In this case, up to 55% of the housing cost gap (0.055 log-points) is explained by differences in LTVs. The result is not surprising, since older mortgages can be refinanced in case of disadvantageous rate deals, and the new interest rate calculations will take into account the reduction in leverage obtained by having repaid a share of the borrowed sum. Since there is usually no interest rate advantage in borrowing at LTVs $\leq 60\%$, refinancing tends to favour the young.

Table 8: Age-specific mortgage costs: recent mortgagers

	ln(opportunity cost)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age 35-44	-0.102***	-0.104***	-0.105***	-0.094***	-0.083***	-0.070***	-0.072***
	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Age 45-64	-0.131***	-0.137***	-0.138***	-0.110***	-0.092***	-0.061***	-0.068***
	(0.014)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)
ln(Mortgage years)		-0.018	-0.017	-0.015	-0.105***	-0.102***	-0.069***
		(0.021)	(0.021)	(0.021)	(0.024)	(0.024)	(0.023)
ln(income/head)			0.010				
			(0.012)				
Age 35-44 × (London+SE)				-0.034		-0.044	-0.042
				(0.028)		(0.028)	(0.028)
Age 45-64 × (London+SE)				-0.081**		-0.093***	-0.087***
				(0.032)		(0.031)	(0.031)
Deposit share					-0.389***	-0.391***	
					(0.032)	(0.032)	
LTV <60%							-0.165***
							(0.020)
LTV 60-70%							-0.090***
							(0.021)
LTV 70-80%							-0.045**
							(0.019)
LTV 80-90%							-0.012
							(0.018)
Observations	5,706	5,706	5,706	5,706	5,687	5,687	5,687
R^2	0.175	0.175	0.176	0.177	0.210	0.212	0.201
Within Adjusted R^2	0.021	0.021	0.021	0.022	0.062	0.064	0.051
Region fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓
Year Purchased fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes:*** ≤ 0.01, ** ≤ 0.05, * ≤ 0.10. . The dependent variable is the imputed opportunity cost of the mortgage, which includes yearly costs plus the shadow cost of the capital used as deposit, which is assumed to be invested in an identical property (and, hence, is revaluated at the same rate of the property's value) with a yearly real yield, on top of re-evaluation reinvested in the property itself, of 2%. The opportunity cost also includes the cost of maintenace, estimated to be 0.895% of the current house value, per year. The cost is then normalised by the initial value of the property. Heteroskedasticity-consistent (HC) standard errors are reported in parentheses.

Table 9: Age-specific mortgage costs: very recent mortgagers

	ln(opportunity cost)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age 35-44	-0.114***	-0.109***	-0.111***	-0.112***	-0.087***	-0.086***	-0.088***
	(0.015)	(0.015)	(0.015)	(0.017)	(0.015)	(0.017)	(0.017)
Age 45-64	-0.123***	-0.105***	-0.107***	-0.100***	-0.057***	-0.045**	-0.056***
	(0.017)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)
ln(Mortgage years)		0.039	0.040	0.040*	-0.035	-0.034	-0.003
		(0.025)	(0.024)	(0.024)	(0.028)	(0.028)	(0.027)
ln(income/head)			0.020				
			(0.017)				
Age 35-44 × (London+SE)				0.008		-0.004	-0.002
				(0.033)		(0.032)	(0.032)
Age 45-64 × (London+SE)				-0.017		-0.035	-0.024
				(0.040)		(0.040)	(0.040)
Deposit share					-0.351***	-0.353***	
					(0.038)	(0.039)	
LTV <60%							-0.150***
							(0.025)
LTV 60-70%							-0.076***
							(0.025)
LTV 70-80%							-0.041*
							(0.024)
LTV 80-90%							-0.015
							(0.023)
Observations	3,179	3,179	3,179	3,179	3,171	3,171	3,171
R^2	0.069	0.070	0.071	0.070	0.108	0.108	0.094
Within Adjusted R^2	0.030	0.031	0.032	0.031	0.071	0.071	0.056
Region fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓
Year Purchased fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes:*** ≤ 0.01, ** ≤ 0.05, * ≤ 0.10. . The dependent variable is the imputed opportunity cost of the mortgage, which includes yearly costs plus the shadow cost of the capital used as deposit, which is assumed to be invested in an identical property (and, hence, is revaluated at the same rate of the property's value) with a yearly real yield, on top of re-evaluation reinvested in the property itself, of 2%. The opportunity cost also includes the cost of maintenace, estimated to be 0.895% of the current house value, per year. The cost is then normalised by the initial value of the property. Heteroskedasticity-consistent (HC) standard errors are reported in parentheses.

A.2.2 Opportunity Cost of Deposits

In the main text, we have built the housing cost index using a 2% yearly dividend for the "indexed" value of the deposit. We now perform a series of robustness checks, showing how the coefficient on the deposit-cost gap changes as we change the expected yearly yield of the deposit from 0% to 3.5%. 0% corresponds to a case where the deposit paid for the house has no opportunity cost, and is invested in an asset which perfectly matches house price inflation. 3.5% corresponds to the average net yield of UK rents, in proportion to the current value of the property, assuming full occupancy throughout all year and no arrears, excluding the present expected value of past and future transaction costs incurred in buying/selling the asset, and the time opportunity cost of managing the few operations that are not delegated to the letting agent. Hence, a 3.5% real return, net of maintenance and operational costs, should be intended as the upper limit. As a reference, notice that the average risk-free real return for the UK was -1.32% for the period of interest (2009-2013), and 1.1% for the period 1999-2013.[31] Hence, our

[31] World Bank statistics for "Real interest rate (%) - United Kingdom".

choice of a baseline of 2% real net return on an alternative investment, in addition to the value growth of the underlying asset, appears to be a reasonable baseline assumption.

To perform the robustness exercise, we repeat the same regression as the one in Column (2) of Table 2 for 0.25% dividend yield increments. We then plot the coefficients and their 90% confidence interval in Figure 10. As expected, we find that the young-old housing cost gap decreases with the opportunity cost of the downpayments, as the old take out mortgages with lower LTV (higher deposit/price ratio) than the young. It ranges from 0.26 log-points when opportunity costs are absent, to 0.09 when we assume them to be 3.5% of the current value of the house, per year.

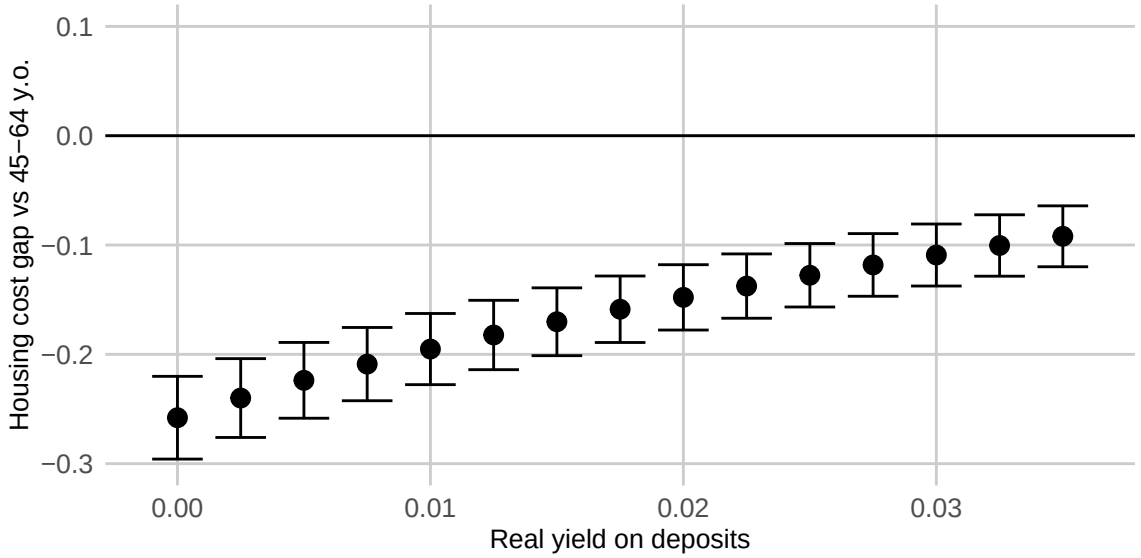


Figure 10: Relative housing cost of 45-64 y.o., by opportunity cost

In Figure 11 we perform a similar exercise, but in relation to the share of the housing-cost gap reported in Figure 10 explained by the LTV-ratio variable. To do so, we re-estimate the regression from Column (4) of Table 2 for 0.25% dividend yield increments. Then, we calculate the ratio between the new 45-64 age coefficients and the ones reported in Figure 10. We call the residual share, defined one minus the coefficient ratio, as the share explained by the mortgage LTVs. We find that, in absence of opportunity costs, over half of the difference in costs would be explained by LTV ratio differences (0.15 log-points). The share is around 33% at our baseline of a 2% opportunity cost (0.05 log-points), and reaches a minimum of 7% (0.006 log-points) at the 3.5% upper limit of the opportunity cost range. Notice that, as discussed in the previous paragraph, the overall housing cost gap of mortgagers remains sizeable (no less than 0.09 log-points at a rate of return of 3.5%), and this exercise only captures the share explained by a linear control for the LTV ratio.

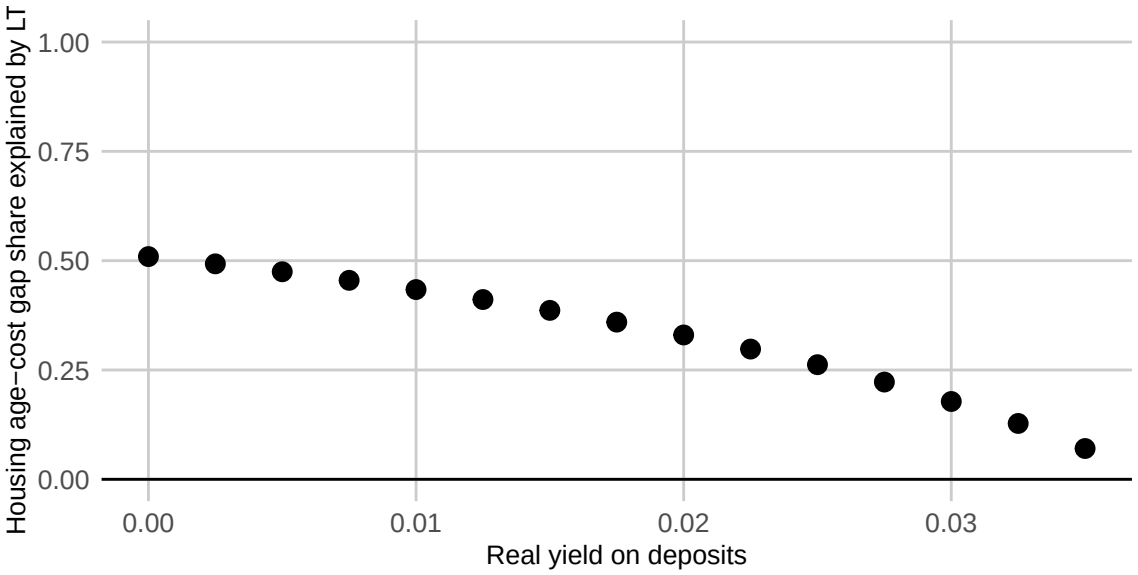


Figure 11: Share of housing cost gap (25-34 vs 45-64 y.o.) explained by LTV ratio

A.2.3 Costs as share of current value

In Table 2, we have used as dependent variable a housing cost index normalised by the initial cost of the house. In this way, we took into account that different individuals could be spending different amounts for different house sizes, or even for identical houses in different cities and neighbourhoods. However, one could argue that this normalisation poses a few issues:

1. If housing has appreciated in real terms since when the deposit was paid, normalising by the initial value may overestimate costs for early buyers (due to the shadow cost of the deposit and maintenance)

2. If housing has depreciated, high yearly mortgage payments should be considered as costly in "current" terms. Hence, normalising by the initial value could lead to underestimate the costs for early buyers

In Table 2 we partially adressed this issue by introducing a year-of-purchase fixed effect. In particular, controlling for year-bought fixed effect takes into account all year-specific house price variations that hit all houses in all locations in the same way. This includes the effect of the consumer price index inflation. Thus, we only need to address the first potential issue: that housing age-costs gaps arise from differences in the appreciation of the house values in an uneven way across space.

We show that our results are robusts to accounting for this effect. In Table 10 we calculate the relative housing costs as a share of house value, as reported in the survey year (rather than its initial value at purchase). We find that the age-gap is somewhat larger but, overall, the coefficients and the role of LTV ratios are in line with our headline results. This is not a surprising result, since the headline results were already reported net of year-bought fixed effects.

Table 10: Age-specific mortgage costs

	ln(opportunity cost)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age 35-44	-0.080***	-0.087***	-0.089***	-0.091***	-0.074***	-0.076***	-0.078***
	(0.011)	(0.012)	(0.012)	(0.013)	(0.011)	(0.013)	(0.013)
Age 45-64	-0.144***	-0.159***	-0.161***	-0.149***	-0.121***	-0.110***	-0.115***
	(0.012)	(0.013)	(0.013)	(0.015)	(0.014)	(0.015)	(0.015)
ln(Mortgage years)		-0.044**	-0.043**	-0.044**	-0.102***	-0.101***	-0.083***
		(0.018)	(0.018)	(0.018)	(0.020)	(0.020)	(0.019)
ln(income/head)			0.025**				
			(0.010)				
Age 35-44 × (London+SE)				0.014		0.007	0.008
				(0.025)		(0.025)	(0.025)
Age 45-64 × (London+SE)				-0.033		-0.036	-0.033
				(0.025)		(0.024)	(0.025)
Deposit share					-0.284***	-0.284***	
					(0.025)	(0.025)	
LTV <60%							-0.133***
							(0.015)
LTV 60-70%							-0.045**
							(0.018)
LTV 70-80%							-0.043***
							(0.016)
LTV 80-90%							-0.017
							(0.014)
Observations	8,107	8,107	8,107	8,107	8,078	8,078	8,078
R^2	0.195	0.196	0.196	0.196	0.213	0.214	0.210
Within Adjusted R^2	0.019	0.020	0.021	0.020	0.040	0.041	0.036
Region fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓
Year Purchased fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes:*** ≤ 0.01, ** ≤ 0.05, * ≤ 0.10. . The dependent variable is the imputed opportunity cost of the mortgage, which includes yearly costs plus the shadow cost of the capital used as deposit, which is assumed to be invested in an identical property (and, hence, is revaluated at the same rate of the property's value) with a yearly real yield, on top of re-evaluation reinvested in the property itself, of 2%. The opportunity cost also includes the cost of maintenace, estimated to be 0.895% of the current house value, per year. The cost is then normalised by the current value of the property. Heteroskedasticity-consistent (HC) standard errors are reported in parentheses.

A.2.4 Renters

We show that the our headline results apply not only to mortgagers, but also to renters. Since the price of a rental properties can be affected by several factors, but we do not know the estimated value of the property, we create a normalised housing index by the number of rooms rented by the household, per household head. This normalisation allows us to account for:

1. Renters who rent individual rooms in a multiple-household occupancy regime, which are counted as separate households
2. Renters sharing their property as a couple, for which a simple costs-over-rooms index (instead of costs over rooms-per-person) would underestimate individual housing costs.

We perform our first estimates using the period 2009-2013, consistently with the main text. To show that our main text facts are not particularly period-dependent, we then show results for the period 1993-2007,

for which we also hold information of the council band of the property (which is based on 1991 market values, and hence can act as a proxy for how expensive, within a region, a property is).

In Table 11 we show that young renters pay 0.256 log-points more than older renters for properties with a similar number of rooms rented per household head, within the same region and with the same number of bedrooms,[32] in the same year. We show that the difference in rental costs falls, but remains significant and comparable to the one in mortgage opportunity costs, once we account for household income per head (-0.209 log-points), or we condition on the number of household heads and adults (approximately -0.09 log-points for 1-head households, and -0.12 for 2-heads households). Finally, we show that the difference appears to be fully explained by the length of time at the property, which is intrinsically connected to age. For example, 25-years old cannot have lived more than 7 years at a property they rent on their own. However, for 50+ years old, it is fairly common to have spent more than 7 years at the same property. Moreover, younger households are more likely to relocate for work, family and education reasons, meaning that their stays tend to be shorter. Once we include the number of years at the property as control, the difference in housing is reduced to virtually zero.

Table 11: Age-specific rental costs

	ln(housing costs per room per adult)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age 35-44	-0.159***	-0.148***	-0.118***	-0.034***	-0.053***	0.0003	-0.013
	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.012)	(0.014)
Age 45-64	-0.256***	-0.209***	-0.143***	-0.093***	-0.122***	0.008	0.030**
	(0.012)	(0.011)	(0.012)	(0.011)	(0.014)	(0.011)	(0.015)
ln(income/head)		0.171***	0.169***	0.153***	0.282***	0.130***	0.244***
		(0.010)	(0.010)	(0.010)	(0.013)	(0.009)	(0.012)
Age 35-44 × (London+SE)			-0.080***	0.025	0.008	0.028	-0.004
			(0.028)	(0.035)	(0.027)	(0.035)	(0.026)
Age 45-64 × (London+SE)			-0.185***	-0.063*	-0.040	-0.054*	-0.043
			(0.026)	(0.033)	(0.030)	(0.032)	(0.028)
Years at property						-0.019***	-0.024***
						(0.0007)	(0.0010)
Observations	25,287	25,287	25,287	11,179	8,906	11,179	8,906
R^2	0.234	0.252	0.255	0.364	0.480	0.419	0.527
Within Adjusted R^2	0.031	0.054	0.057	0.074	0.129	0.153	0.208
Region fixed effects	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓
Bedrooms fixed effects	✓	✓	✓	✓	✓	✓	✓
Household size	All	All	All	1	2	1	2

Notes:*** ≤ 0.01, ** ≤ 0.05, * ≤ 0.10. . The dependent variable is the housing cost faced by renters, normalised by the number of rooms per household adult. The regression is estimated over the period 2009-2013, consistently with the headline results on mortgages. Heteroskedasticity-consistent (HC) standard errors are reported in parentheses.

In Table 12 we perform an identical analysis for the period 1993-2007. We do so for two reasons. First, we show that the housing cost gap on renters is not specific of a particular time period. Using the same controls as specification (1) in Table 11, we find similar age gaps for the 1993-2007 period. Second, the 1993-2007 editions of the English Housing Survey contained a question regarding the Council Tax Band of the property. Council taxes are paid by occupiers, rather than owners, and are based on "bands"

[32]We control for the number of bedrooms, besides normalising for the number of rooms rented by the renter, to account for how certain rooms may be more "valuable" or expensive to furnish than others and, hence, may carry a larger rental premium.

assigned according to the market value of the property in 1991.[33] While the actual tax for the "central" band is chosen by the local council, its value classification in 1991 terms is given by law. Hence, the council tax band represents a good proxy for the value of the property. In column (2), controlling for the band increases the estimated housing gap by 0.03 log-points in favour of the old. The remaining columns replicate the specifications of Table 11. Again, introducing a "years at property" control eliminates all statistically significant differences in rental price gaps.

Table 12: Age-specific rental costs, 1993-2007

	ln(housing costs per room per adult)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age 35-44	-0.180***	-0.191***	-0.139***	-0.114***	-0.020	-0.057**	0.017	0.010
	(0.015)	(0.016)	(0.017)	(0.018)	(0.021)	(0.024)	(0.020)	(0.022)
Age 45-64	-0.315***	-0.346***	-0.275***	-0.240***	-0.142***	-0.232***	-0.025	0.009
	(0.015)	(0.016)	(0.017)	(0.019)	(0.022)	(0.026)	(0.022)	(0.025)
ln(income/head)			0.062***	0.061***	0.078***	0.075***	0.058***	0.051***
			(0.010)	(0.010)	(0.011)	(0.014)	(0.011)	(0.013)
Age 35-44 × (London+SE)				-0.070*	-0.010	0.006	-0.012	-0.025
				(0.037)	(0.046)	(0.043)	(0.045)	(0.041)
Age 45-64 × (London+SE)				-0.103***	-0.007	0.001	0.057	-0.038
				(0.038)	(0.045)	(0.052)	(0.043)	(0.048)
Years at property							-0.031***	-0.043***
							(0.002)	(0.002)
Observations	16,360	12,974	10,837	10,837	5,170	5,285	5,169	5,279
R^2	0.311	0.347	0.337	0.337	0.374	0.455	0.422	0.521
Within Adjusted R^2	0.030	0.040	0.033	0.034	0.025	0.031	0.100	0.149
Region fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Bedrooms fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Council Tax Band fixed effects		✓	✓	✓	✓	✓	✓	✓
Household size	All	All	All	All	1	2	1	2

Notes:*** ≤ 0.01, ** ≤ 0.05, * ≤ 0.10. . The dependent variable is the housing cost faced by renters, normalised by the number of rooms per household adult. The regression is estimated over the period 2009-2013, consistently with the headline results on mortgages. Heteroskedasticity-consistent (HC) standard errors are reported in parentheses.

A.2.5 Downpayments and Mortgage Length

Figure 12 shows that the young tend to have higher LTV ratios than the old. The young provide an average downpayment of 18% of the total equity, against an average of 28% for the 45-64 years old. That is, they borrow 14% more equity than the average old to buy equally-priced houses. On the other hand, the young tend to take longer mortgages. The average mortgage length at origination for a 25-34 years old is 26 years, compared to 20 years for a 45-64 years old. That is, young households' mortgages are 30% longer.

While this could be a symptom of long-term credit being harder to get for older households (**?**), notice how the young still pay a higher opportunity cost of housing due to higher LTV (smaller deposits) commanding larger interest rates. Even more, the average payment is much more sensitive to interest rates for long-duration loans, rather than short-term ones, leading to higher risk. Finally, longer mortgages also lead to repaying a smaller share of equity every period, meaning that present value of all future interest payments is higher than what we capture using the flow-cost of housing.

[33] New dwellings built after 1991 are assigned to a band by imputing their 1991 value, based on similar nearby properties and the price growth of the area.

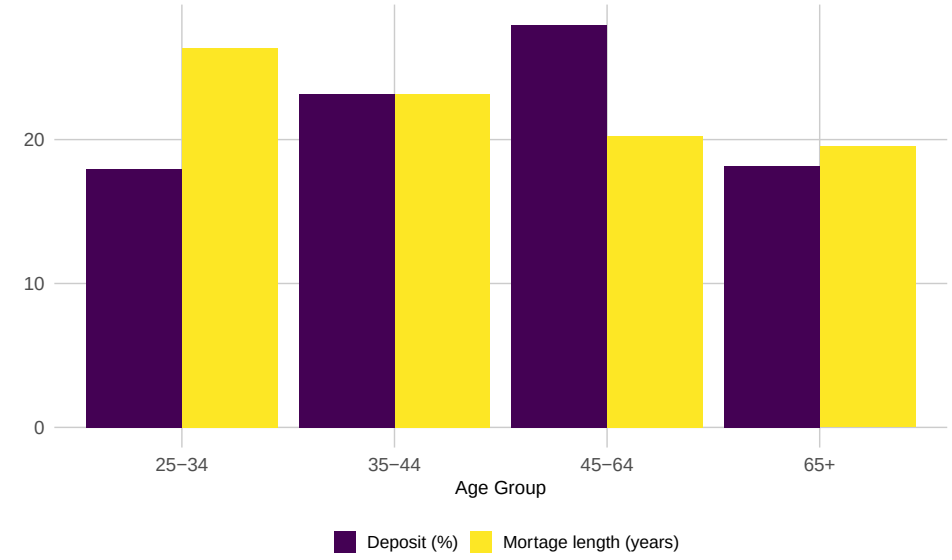
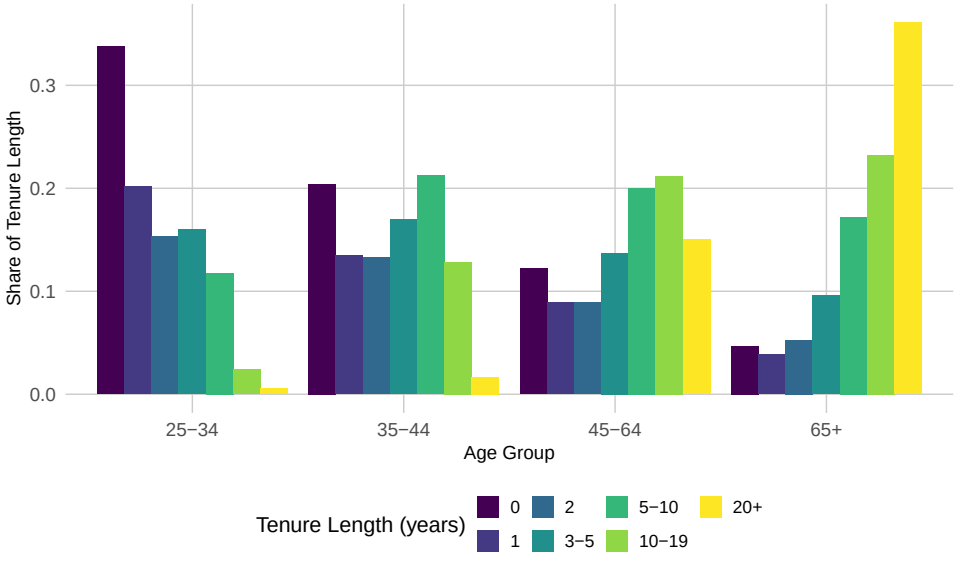


Figure 12: Deposit (as share of equity) and Mortgage Length, by age group

A.2.6 Years at address

We showed, in Table 11, that the differences in rental costs can be explained by controlling for the years spent at the address. In this section, we show that there are clear age-related patterns in the length of time at an address for renters. Figure 13 shows how one-third of 25-34 years old has been living at the address where they were interviewed for less than a year, and over 50% had lived at the property less than two years. Conversely, two-thirds of 45-64 years old renters had lived at the property for more than five years. The inversion of the pattern is already clear in the 35-44 age group, where over one-third of households has been at the same address for over 5 years, a slightly higher proportion than individuals who had been at the same address less than two years at the moment of the interview.

Figure 13: Years at address, by age - Renters



A.3 Ageing and Productivity: Additional Robustness

Figure 14: Ageing and Productivity, per year, 2002-2019



The figure shows annual changes in ageing from year $t-1$ to t against the log of value added per worker in year $t-1$, again in deviations from the national mean, for each year from 2003 to 2019. TTWAs are divided by population density: yellow circles indicate areas below the median density, while blue diamonds represent areas above it. Separate linear regressions are shown for each subgroup, with 95% confidence intervals.

Table 13 complements the main results in Table **??** by introducing a range of robustness checks. Column (1) reproduces the benchmark specification discussed in the main text. Column (2) presents an unweighted regression to assess the influence of population weighting. Columns (3) and (4) test sensitivity to the choice of spatial correlation structure by adjusting the Conley standard error radius to 30 km and 100 km, respectively. Column (5) introduces a higher-order lag structure, using a third-order lag operator for ageing. Finally, Column (6) adopts an alternative measure of ageing, defined as the year-on-year change in the average age of the working-age population. Across all these specifications, the association between local productivity (value added per worker) and ageing remains positive and statistically significant, underscoring the robustness of the main findings.

Table 13: Ageing and Value Added per worker: Additional Robustness

	Ageing					Mean Age
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.007	0.020*	0.007	0.007	0.007	0.004
	(0.006)	(0.012)	(0.005)	(0.008)	(0.007)	(0.003)
Value Added (lagged)	0.128***	0.142***	0.128***	0.128***	0.155***	0.083***
	(0.016)	(0.032)	(0.020)	(0.017)	(0.016)	(0.009)
Density (lagged)	-0.0010***	-0.017*	-0.010*	-0.010***	0.001	0.015***
	(0.003)	(0.007)	(0.005)	(0.003)	(0.004)	(0.005)
Change HP (lagged)	0.502***	0.469**	0.502**	0.502***	1.095***	0.134
	(0.188)	(0.168)	(0.184)	(0.141)	(0.314)	(0.117)
Ageing (lagged)	0.768***	0.584***	0.768***	0.768***	0.839***	1.024***
	(0.036)	(0.031)	(0.037)	(0.016)	(0.030)	(0.004)
Observations	2,768	2,768	2,768	2,768	2,595	2,768
R^2 (adj.)	0.625	0.608	0.625	0.625	0.608	0.998
Weights	WA pop	-	WA pop	WA pop	WA pop	WA pop
Lags	1	1	1	1	3	1
Radius Conley SE	60	60	30	100	60	60

Notes:*** ≤ 0.01, ** ≤ 0.05, * ≤ 0.10. Column (1) reproduces the baseline from Table **??** using the gap in shares of older (50–64) vs. younger (25–34) workers. Columns (2)–(5) test robustness using the same dependent variable but different lag structures and Conley bandwidths. Column (6) uses the change in mean age of the working-age population. All regressions use Conley standard errors. Populations used for weights are noted in the table.

B Additional material for Section 3

B.1 Derivation of the old and young workers' location choice

Old Workers Under Assumption S.3, the maximum idiosyncratic component is larger than the maximum difference in non-idiosyncratic components of old age utility from living in c or r, since $\sigma > |\Delta\mu^Y + \Delta Z| + |\Delta\mu^O + \Delta Z(1+e+e_C)| + |Z_R e_C| > |\Delta\mu^O + \Delta Z(1+e+e_C)|$. Hence, there is no mass of individuals that, for a given skill level, would always choose j over $\bar{j}$ for all possible values of $(\varepsilon_j, \varepsilon_{\bar{j}})$. An old individual with a skill level $s^O(j^Y)$ chooses location j over $\bar{j}$ when its utility in j is higher, i.e.:

$$\mu_j^O + w_{j,t} s^O(j^Y) - (1-\psi)P_{j,t} + \tau_t + \varepsilon_{j,t} > \mu_{\bar{j}}^O + w_{\bar{j},t} s^O(j^Y) - (1-\psi)P_{\bar{j},t} + \tau_t + \varepsilon_{\bar{j},t},$$

or, equivalently,

$$\varepsilon_{\bar{j},t} - \varepsilon_{j,t} < \mu_j^O - \mu_{\bar{j}}^O + (w_{j,t} - w_{\bar{j},t}) s^O(j^Y) - (1-\psi)(P_{j,t} - P_{\bar{j},t}).$$

Due to assumption S.2, we know that $\varepsilon_{\bar{j},t} - \varepsilon_{j,t} \sim U(-\sigma, \sigma)$. Hence, denoting with $\phi_{j,t}^O(s^O(j^Y))$ the probability that an old individual with skill $s^O(j^Y)$ chooses location j, we know that this is equal to the probability that $\varepsilon_{\bar{j},t} - \varepsilon_{j,t} < \mu_j^O - \mu_{\bar{j}}^O + (w_{j,t} - w_{\bar{j},t}) s^O(j^Y) - (1-\psi)(P_{j,t} - P_{\bar{j},t})$, which is:

$$\begin{aligned}
\phi_{j,t}^O(s^O(j^Y)) &= \frac{\mu_j^O - \mu_{\bar{j}}^O + (w_{j,t} - w_{\bar{j},t}) s^O(j^Y) - (1-\psi)(P_{j,t} - P_{\bar{j},t}) + \sigma}{2\sigma} \\
&= \frac{1}{2} + \frac{1}{2\sigma}\left[\left(\mu_j^O - \mu_{\bar{j}}^O\right) + (w_{j,t} - w_{\bar{j},t}) s^O(j^Y) - (1-\psi)(P_{j,t} - P_{\bar{j},t})\right], \quad \text{with } j \neq \bar{j}.
\end{aligned}$$

Young Workers Recall the young worker's problem:

$$V^Y(s^Y,\varepsilon_t) = \max_j \left\{\mu_j^Y + w_{j,t}s^y + \tau_t - P_{j,t} + \varepsilon_{j,t} + \mathbb{E}[V^O(s^O(j),\varepsilon_{t+1})]\right\};$$
$$s^O(j) = s^Y(1+e+e_j),$$

When comparing the available locations today, a young worker considers the difference in instantaneous utility and in expected utility.

$$\varepsilon_{\bar{j}} - \varepsilon_j \leq (\mu_j^Y + w_{j,t}s^y + \tau_t - P_{j,t}) - (\mu_{\bar{j}}^Y + w_{\bar{j},t}s^y + \tau_t - P_{\bar{j},t}) + E_t[V^O(s^O(j),\varepsilon_{t+1})] - E_t[V^O(s^O(\bar{j}),\varepsilon_{t+1})]. \tag{15}$$

To compute the final probability, we need to develop the difference in expecetd utility. Recall that, due to Assumption S.4,

$$E_t[V^O(s^O(j),\varepsilon_{t+1})] = \sum_{j'} \phi^O_{j',t+1}(S^O(j))\left(\mu_{j'}^O + w_{j',t+1}(s^O(j^Y)) - (1-\psi)P_{j,t+1} + \tau_t\right).$$

We are interested in the difference of expected values next period utility between choosing today either j or $\bar{j}$, i.e.: $E_t[V^O(s^O(j),\varepsilon_{t+1})] - E_t[V^O(s^O(\bar{j}),\varepsilon_{t+1})]$.

Hence, the difference expected value of tomorrow's utility from choosing today either j or $\bar{j}$ is:

$$\sum_{j'}\left[\phi^O_{j',t+1}\left(s^O(j)\right)\left(\mu_{j'}^O + w_{j',t+1}s^Y(1+e+e_j) - (1-\psi)P_{j',t+1} + \tau_t\right)\right]$$
$$-\sum_{j'}\left[\phi^O_{j',t+1}\left(s^O(\bar{j})\right)\left(\mu_{j'}^O + w_{j',t+1}s^Y(1+e+e_{\bar{j}}) - (1-\psi)P_{j',t+1} + \tau_t\right)\right].$$

Consider the probabilities of choosing j' when old $\left(\phi^O_{j',t+1}\left(s^O(j^Y)\right)\right)$ in the case when j^Y is either j or $\bar{j}$. From equation (**??**), the only term which differes once j^Y changes is the location specific human capital (e_{j^Y}). Therefore, we can write:

$$\phi^O_{j',t+1}\left(s^O(j^Y=j)\right) = \phi^O_{j',t+1}\left(s^O(j^Y=\bar{j})\right) + \frac{w_{j',t+1}s^Y(e_j - e_{\bar{j}})}{2\sigma}.$$

Hence, collecting first the differences in income ($w_{j',t+1}e_{j^Y}$) across $j^Y \in \{j,\bar{j}\}$ and then the differences in location choice probabilities, we can write the difference of expected values as

$$s^Y\sum_{j'} w_{j',t+1}\left(\phi^O_{j',t+1}(s^O(j))e_j - \phi^O_{j't+1}(s^O(\bar{j}))e_{\bar{j}}\right) + \frac{s^Y(e_j-e_{\bar{j}})}{2\sigma}\sum_{j'} w_{j',t+1}\left(\mu_{j'}^O + w_{j',t+1}(1+e) - (1-\psi)P_{j',t+1} + \tau_t\right).$$

Substituting this difference into equation (15) and following the same steps for the old worker problem, Assumption S.2 directly implies:

$$\begin{aligned}\phi^Y_{j,t}(s^Y) =& \frac{1}{2} + \frac{1}{2\sigma}\left[\left(\mu_j^Y - \mu_{\bar{j}}^Y\right) + (w_{j,t} - w_{\bar{j},t})s^Y - (P_{j,t} - P_{\bar{j},t})\right] \\ &+ \frac{1}{2\sigma}\left[\mathbb{E}[V^O(s^O(j),\varepsilon_{t+1})] - \mathbb{E}[V^O(s^O(\bar{j}),\varepsilon_{t+1})]\right]\end{aligned} \tag{16}$$

Assumption S.3 ensures that ϕ_j^Y can be interpreted as a probability, as it is always bounded between 0 and 1 due to the maximum idiosyncratic component always dominating the non-idiosyncratic component of each location, calculated at youth (thus accounting for both young-age utility and its continuation

value).

B.2 Definition of a Steady State equilibrium

Definition 1 (Steady state)**.** A steady state (or stationary equilibrium) is a set

$$\left(\tilde{\phi}_j^Y(s^Y), \tilde{\phi}_j^O\left(s^O(j^Y=j)\right), \tilde{\phi}_j^O\left((s^O(j^Y=\bar{j})\right), \tilde{L}_j, \tilde{\tau}, \tilde{w}_j, \tilde{P}_j\right), \forall j \in \{C,R\},$$

such that the sequences

$$\left\{\phi_{j,t}^Y(s^Y), \phi_{j,t}^O\left(s^O(j^Y=j)\right), \phi_{j,t}^O\left((s^O(j^Y=\bar{j})\right), L_{j,t}, \tau_t, w_{j,t}, P_{j,t}\right\}_{t=0}^{\infty}, \forall j,$$

defined by

$$\begin{aligned}
\phi_{j,t}^Y(s^Y) &= \tilde{\phi}_j^Y(s^Y), \\
\phi_{j,t}^O\left(s^O(j^Y=j)\right) &= \tilde{\phi}_j^O\left(s^O(j^Y=j)\right), \\
\phi_{j,t}^O\left((s^O(j^Y=\bar{j})\right) &= \tilde{\phi}_j^O\left((s^O(j^Y=\bar{j})\right), \\
L_{j,t} &= \tilde{L}_j, \\
\tau_t &= \tilde{\tau}, \\
w_{j,t} &= \tilde{w}_j, \\
P_{j,t} &= \tilde{P}_j,
\end{aligned}$$

are an equilibrium for the initial condition $\phi_{j,-1}^Y\left(s^Y\right) = \tilde{\phi}_j^Y(s^Y)$.

In other words, a steady state is an equilibrium for which the allocations are constant over time.

B.3 Definition of Sequential Market Equilibrium

Definition 2 (Sequential Market Equilibrium)**.** Given an initial skill distribution of young workers at time 0 in each location j, $\phi_{j,0}^Y\left(s^Y\right)$, a sequential market equilibrium is: masses of young workers in location j, $\{\phi_{j,t}^Y(s^Y)\}_{t=0}^{\infty}$, of old workers with skill $s^O(j^Y=j)$, $\{\phi_{j,t}^O\left(s^O(j^Y=j)\right)\}_{t=0}^{\infty}$ and of old workers with skill $s^O(j^Y=\bar{j})$, $\{\phi_{j,t}^O\left((s^O(j^Y=\bar{j})\right)\}_{t=0}^{\infty}$ and with $\bar{j} \neq j$; a demand for labor, $\{L_{j,t}\}_{t=0}^{\infty}$; transfers $\{\tau_t\}_{t=0}^{\infty}$; wages $\{w_{j,t}\}_{t=0}^{\infty}$; house prices $\{P_{j,t}\}_{t=0}^{\infty}$, $\forall j \in \{C,R\}$, such that:

1. Given the sequence of wages and housing prices, $(w_{j,t}, P_{j,t})$ for all t and j, $\phi_{j,t}^Y(s^Y)$, $\phi_{j,t}^O\left(s^O(j^Y=j)\right)$, and $\phi_{j,t}^O\left(s^O(j^Y=\bar{j})\right)$ are consistent with the maximizing behavior of the workers, described by equations (**??**) and (16).

2. For all t, given $w_{j,t}$, the firm maximizes profits, i.e. $w_{j,t} = Z_j$.

3. For all t, the housing market clears in each location j:

$$\phi_{j,t}^Y(s^Y) + \phi_{j,t}^O\left(s^O(j^Y=j)\right)\phi_{j,t-1}^Y(s^Y) + \phi_{j,t}^O(s^O(j^Y=\bar{j})(1-\phi_{j,t-1}^Y) = n_j.$$

4. For all $t > 0$, the revenue of the housing sector is uniformily distributed to each individual in the economy, i.e.:

$$\tau_t = \frac{1}{2}\left(n_C P_{C,t} + n_R P_{R,t}\right).$$

Notice that the initial condition necessary to determine the sequential competitive equilibrium is the origin of the old workers in period 1. In fact, $\phi^Y_{j,0}\left(s^Y\right)$ determines the mass of old workers with skill $1+e+e_j$ at time 1, while $1-\phi^Y_{j,0}\left(s^Y\right)$ determines the mass of old workers with skill $1+e+e_{\tilde{j}}$. With this information, the skill distribution at time 1 in the economy is fully characterized.

Also, because of the assumption of fixed housing supply, the equilibrium determines only the difference in location price $\Delta P_t = P_{j,t} - P_{\tilde{j},t}$. This is not an issue for this section as we are interested in comparing location affordability across generations. Therefore, for the rest of this section, we assume without loss of generality that $P_{R,t} = 0$, $\forall t$, so that the $P_{C,t}$ should be interpreted as the additional cost of the city compared to the rural area.[34]

Since firms are indifferent between demanding any amount of labor for $w_j = Z_j$, and due to assumptions S.1 and S.3, we know that there is always a positive supply of labor in each location, then both the city and rural labor markets clear at $w_j = Z_j$. Finally, the tradeable goods market also clears since all other markets clear.

B.4 Characterization of the Sequential Market Equilibrium

Assume Assumptions S.1, S.2, S.3 and S.4 hold. Also, assume that $n_C = n_R = 1$, and $s^Y = 1$. Any interior solution of the sequential market equilibrium, such that $\phi^Y_{C,t} \in (0,1)$, satisfies the following system of equations:

$$\begin{cases} \phi^Y_{C,t+1} + \left[\frac{\sigma(2-\psi)}{(1-\psi)\Delta Z e_C} + \frac{\Delta Z e_C}{2\sigma}\right]\phi^Y_{C,t} + \frac{1}{2(1-\psi)}\phi^Y_{C,t-1} = \\ \qquad \frac{1}{2(1-\psi)\Delta Z e_C}\left[\sigma(2-\psi) + (1-\psi)\Delta\mu^Y - \Delta\mu^O - \psi\Delta Z \right. \\ \qquad\qquad \left. + \frac{5}{2}(1-\psi)\Delta Z e_c - \Delta Z e + (1-\psi)\frac{\Delta Z^2 e_C^2}{2\sigma}\right] \\ \phi^O_{C,t}(s^O(j^Y)) = 1 + \frac{\Delta Z e_{j^Y}}{2\sigma} - \phi^Y_{C,t} - \frac{\Delta Z e_C}{2\sigma}\phi^Y_{C,t-1}, \\ P_{C,t} = \frac{1}{1-\psi}\left[2\sigma\phi^Y_{C,t} + \Delta Z e_C \phi^Y_{C,t-1} + \Delta\mu^O - \sigma + \Delta Z(1+e)\right], \\ P_{C,t} \in \left(\Delta\mu^O + \Delta Z(1+e+e_C\phi^Y_{C,t-1}) - \sigma, \Delta\mu^O + \Delta Z(1+e+e_C\phi^Y_{C,t-1}) + \sigma\right), \\ P_{C,t} < 2Z_C. \end{cases} \tag{17}$$

To derive the conditions above, let us start from the joint dynamic of the share of young in the city (equations 16), the share of old in the city for any level of skill $s^O(j^y)$ (equation **??**), and the market clearing condition for the city when $n_C = n_R = 1$. The final inequality, $P_{C,t} < 2Z_C$, provides a sufficient condition to ensure that Assumption S.1 is satisfied. Also, from the firms' problem in equilibrium we have: $w_{jt} = Z_{jt}$. We have:

$$\begin{cases} \phi^Y_{C,t} & = \frac{1}{2} + \frac{1}{2\sigma}(\Delta\mu^Y + \Delta Z - \Delta P_t) + \frac{1}{4\sigma^2}\Delta Z e_C\left(\Delta\mu^O + \Delta Z(1+e) - (1-\psi)\Delta P_{t+1}\right) \\ & \quad + \frac{1}{2\sigma}\left(Z_R(1-\phi^O_{C,t+1}(s^O(C)))e_C + (Z_R+\Delta Z)\phi^O_{C,t+1}(s^O(C))e_C\right), \\ \phi^O_{C,t}(s^O(j^Y)) & = \frac{1}{2} + \frac{1}{2\sigma}\left[\Delta\mu^O + \Delta Z s^O(j^Y) - (1-\psi)\Delta P_t\right], \\ 1 & = \phi^Y_{C,t} + \phi^O_{C,t}(s^O(C))\phi^Y_{C,t} + \phi^O_{C,t}(s^O(R))(1-\phi^Y_{C,t}) \\ P_{C,t} & < 2Z_C, \quad \forall t. \end{cases}$$

By substituting the second expression into the third, for $s^O(j^Y) = (1+e+e_{j^Y})$, we can write the market

[34] The quantitative model in section X relaxes the assumption of perfectly inelastic housing supply.

clearing equation in terms of the share of young in the city at $t-1$ and t, and prices at t:

$$\begin{aligned}1-\phi^Y_{C,t} = \phi^O_{C,t} &= \phi^O_{C,t}(s^O(C))\phi^Y_{C,t-1} + \phi^O_{C,t}(s^O(R))(1-\phi^Y_{C,t-1}) \\ &= \left(\phi^O_{C,t}(s^O(R)) + \frac{\Delta Z e_C}{2\sigma}\right)\phi^Y_{C,t-1} + \phi^O_{C,t}(s^O(R))(1-\phi^Y_{C,t-1}) \\ &= \frac{1}{2} + \frac{1}{2\sigma}\left(\Delta\mu^O + \Delta Z(1+e) - (1-\psi)\Delta P_t\right) + \frac{\Delta Z e_C}{2\sigma}\phi^Y_{C,t-1}.\end{aligned}$$

Hence, the market clearing condition yields an expression for the market-clearing sequence of prices $\{P_{C,t}\}_{t\geq 1}$:

$$(1-\psi)P_{C,t} = \left(2\sigma\phi^Y_{C,t} + \Delta Z e_C \phi^Y_{C,t-1}\right) + \Delta\mu^O - \sigma + \Delta Z(1+e), \quad \forall t.$$

Substituting the into the equations for shares of young and old in the city, we obtain the following system of equations:

$$\begin{cases} \phi^Y_{C,t+1} + \frac{\sigma}{\Delta Z e_C}\left(\frac{1+(1-\psi)}{(1-\psi)} + \frac{\Delta Z^2 e_C^2}{2\sigma^2}\right)\phi^Y_{C,t} + \frac{1}{2(1-\psi)}\phi^Y_{C,t-1} = \\ \quad \frac{1}{2(1-\psi)\Delta Z e_C}\left[\sigma(1+(1-\psi)) + ((1-\psi)\Delta\mu^Y - \Delta\mu^O) + ((1-\psi)\Delta Z e_c - \Delta Z e) + \Delta Z((1-\psi)-1) + \frac{3}{2}(1-\psi)\Delta Z e_c - \right. \\ \phi^O_{C,t}(s^O(j^Y)) = 1 + \frac{1}{2\sigma}\Delta Z e_{j^Y} - \left(\phi^Y_{C,t} + \frac{\Delta Z e_C}{2\sigma}\phi^Y_{C,t-1}\right), \\ P_{C,t} = \frac{1}{(1-\psi)}\left[\left(2\sigma\phi^Y_{C,t} + \Delta Z e_C \phi^Y_{C,t-1}\right) + \Delta\mu^O - \sigma + \Delta Z(1+e)\right], \quad \forall t. \\ P_{C,t} \in \left(\Delta\mu^O + \Delta Z(1+e+e_C\phi^Y_{C,t-1}) - \sigma, \Delta\mu^O + \Delta Z(1+e+e_C\phi^Y_{C,t-1}) + \sigma\right), \quad \forall t \\ P_{C,t} < 2Z_C, \quad \forall t. \end{cases}$$

The forth expression- a boundary condition for the city price- follows from the fact that the sequence of prices needs to guarantee that $\phi^Y_{C,t} \in (0,1)$ for any initial condition, as the dynamic equations have been derived under the assumption of an internal solution for $\phi^O_{j,t}(\cdot)$ and $\phi^Y_{j,t}(\cdot)$.Assumption S.3 guarantees that this condition holds. Simplifying further the algebra gives the dynamic equations in (17).

Proposition 3 assures that the equilibrium will be always interior.

B.5 Properties of the Sequential Market equilibrium

The sequantil market equilibrium defined in Appendix B.3 and characterized in Appendix B.4 has the following properties.

Proposition 3. *Assume Assumptions S.1, S.2, S.3 and S.4 hold. Also, assume that $n_C = n_R = 1$, and $s^Y = 1$. Then:*

P.1 *The system of equations that defines the sequential market equilibrium (see equations (17)) has*

one unique steady state $\tilde{\phi}_C^Y$, which is given by:[35]

$$\tilde{\phi}_C^Y = \frac{1}{2} + \frac{\overbrace{(1-\psi)\Delta\mu^Y - \Delta\mu^O}^{\text{Net amenity pull}}}{(4-2\psi)\sigma + (3-2\psi)\Delta Z e_C + (1-\psi)\frac{(\Delta Z e_C)^2}{\sigma}} + \underbrace{\frac{\overbrace{(1-\psi)(Z_R e_C - e\Delta Z)}^{\substack{\text{Human capital considerations} \\ \text{(scaled by financial access)}}}}{(4-2\psi)\sigma + (3-2\psi)\Delta Z e_C + (1-\psi)\frac{(\Delta Z e_C)^2}{\sigma}}}_{\text{Spatial concentration: wage and return to experience}} +$$

$$- \frac{\overbrace{\psi\Delta Z(1+e+\frac{e_C}{2})}^{\substack{\text{Shadow cost of constraint:} \\ \text{Inaccessible wage premium}}}}{(4-2\psi)\sigma + (3-2\psi)\Delta Z e_C + (1-\psi)\frac{(\Delta Z e_C)^2}{\sigma}}.$$

P.2 *The steady state is stable, i.e. for any $\phi_{C,t}^Y \neq \tilde{\phi}_C^Y$, $\lim_{s\to\infty} \phi_{C,t+s}^Y = \tilde{\phi}_C^Y$.*

P.3 *The convergence to the steady state is cyclical. If $\phi_{C,t}^Y - \tilde{\phi}_C^Y > 0$, then $\phi_{C,t+1}^Y - \tilde{\phi}_C^Y < 0$ and viceversa, with the sequence $\left\{|\phi_{C,t+1}^Y - \tilde{\phi}_C^Y|\right\}$ being monotonically decreasing and converging to zero.*

Proof.
Proof of P.1. First, let us rewrite the first equation of the system in equation (17) as:

$$\phi_{C,t+2}^Y - 2\alpha\phi_{C,t+1}^Y + b\phi_{C,t}^Y = \gamma, \tag{18}$$

with
$$\alpha = -\frac{1}{2}\left[\frac{\sigma(2-\psi)}{(1-\psi)\Delta Z e_C} + \frac{\Delta Z e_C}{2\sigma}\right],$$
$$b = \frac{1}{2\psi},$$
$$\gamma = \frac{1}{2(1-\psi)\Delta Z e_C}\left[\sigma(2-\psi) + (1-\psi)\Delta\mu^Y - \Delta\mu^O - \psi\Delta Z + \frac{5}{2}(1-\psi)\Delta Z e_c - \Delta Z e + (1-\psi)\frac{\Delta Z^2 e_C^2}{2\sigma}\right].$$

The steady state of this difference equation is:

$$\tilde{\phi}_C^Y = \frac{\gamma}{1-2\alpha+(1-\psi)}$$
$$= \frac{\frac{1}{2(1-\psi)\Delta Z e_C}\left[\sigma(2-\psi) + (1-\psi)\Delta\mu^Y - \Delta\mu^O + (1-\psi)\Delta Z e_c - \Delta Z e + \Delta Z((-\psi) + \frac{3}{2}(1-\psi)\Delta Z e_c + (1-\psi)\frac{\Delta Z^2 e_C^2}{2\sigma}\right]}{1 + \frac{1}{2(1-\psi)} + \frac{\sigma}{\Delta Z e_C}\left(\frac{2-\psi}{(1-\psi)} + \frac{\Delta Z^2 e_C^2}{2\sigma^2}\right)}.$$

Multiplying above and below by $2(1-\psi)\Delta Z e_C$, we obtain:

[35] Appendix**??** derives the steady states for the other endogenous variables.

$$\begin{aligned}\tilde{\phi}_C^Y =& \frac{\sigma(2-\psi)+(1-\psi)\Delta\mu^Y-\Delta\mu^O+(1-\psi)\Delta Ze_c-\Delta Ze-\psi\Delta Z+\frac{3}{2}(1-\psi)\Delta Ze_c+(1-\psi)\frac{\Delta Z^2e_C^2}{2\sigma}}{2\sigma(2-\psi)+\Delta Ze_C\left(3-2\psi+(1-\psi)\frac{\Delta Ze_C}{\sigma}\right)} \\ =& \frac{1}{2}+\frac{(1-\psi)\Delta\mu^Y-\Delta\mu^O}{2(2-\psi)\sigma+\Delta Ze_C(3-2\psi+(1-\psi)\frac{\Delta Ze_C}{\sigma})}+\frac{(1-\psi)Z_Re_C-e\Delta Z}{2(2-\psi)\sigma+\Delta Ze_C(3-2\psi+(1-\psi)\frac{\Delta Ze_C}{\sigma})} \\ & -\psi\frac{\Delta Z(1+\frac{e_C}{2})}{2(2-\psi)\sigma+\Delta Ze_C(3-2\psi+(1-\psi)\frac{\Delta Ze_C}{\sigma})}.\end{aligned}$$

Including the steady state equations for the old allocation in the locations R and C, we have the fully characterised steady state allocation of the market:

$$\begin{aligned}\tilde{\phi}_C^O(s^O(C)) &= \left(1-\tilde{\phi}_C^Y\right)\left(1+\frac{\Delta Ze_C}{2\sigma}\right), \\ \tilde{\phi}_C^O(s^O(R)) &= \left(1-\tilde{\phi}_C^Y\right)\left(1+\frac{\Delta Ze_C}{2\sigma}\right)-\frac{\Delta Ze_C}{2\sigma}.\end{aligned}$$

Proof of P.2. Let us rewrite the second-order difference equation in (18) as a system of first-order difference equations. Denote with $x_t=\phi_{C,t}^Y$, with $y_t=\phi_{C,t-1}^Y$, and with $\nu=\frac{\sigma}{\Delta Ze_C}$. Then, equation (18) is equivalent to:

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{bmatrix}x_{t+1}\\ y_{t+1}\end{bmatrix}=\begin{pmatrix}2\alpha & -(1-\psi)\\ 1 & 0\end{pmatrix}\begin{bmatrix}x_t\\ y_t\end{bmatrix}+\begin{pmatrix}\gamma\\ 0\end{pmatrix}.$$

The system has eigenvalues $\lambda=\alpha\pm\sqrt{\alpha^2-(1-\psi)}$, with solutions:

$$\begin{cases}\lambda_1=-\left(\nu+\frac{1}{4}\frac{1}{\nu}\right)+\sqrt{\nu^2+\frac{1}{16}\frac{1}{\nu^2}};\\ \lambda_2=-\left(\nu+\frac{1}{4}\frac{1}{\nu}\right)-\sqrt{\nu^2+\frac{1}{16}\frac{1}{\nu^2}}.\end{cases}$$

Due to Assumption S.3, $\sigma>\max\{|\Delta Ze_C|,|(1-\psi)\Delta Ze_C|\}$. Hence, $\nu>1$ or $\nu<-1$. We will now consider these two possible cases.

First, consider the case in which $\nu>1$. That implies $-\left(\nu+\frac{1}{4}\frac{1}{\nu}\right)<-1$. Hence, $|\lambda_2|>1$. We now demonstrate that $|\lambda_1|<1$. Define $\rho=\nu+\frac{1}{4}\frac{1}{\nu}$. Then,

$$|\lambda_1|<1 \qquad \text{if} \quad \left\{-\rho+\sqrt{\rho^2-\frac{1}{2}}<1 \qquad \wedge \qquad \rho-\sqrt{\rho^2-\frac{1}{2}}>-1\right\}.$$

This condition is always satisfied for $\rho>-\frac{1}{4}$. However, since $\rho>1$, this means that $|\lambda_1|<1$ is always satisfied for $\nu>1$. Then, in this case, one eigenvalue of the system lies insider the unit circle and the other lies outside.

Second, consider the case in which $\nu<-1$. This implies $-\left(\nu+\frac{1}{4}\frac{1}{\nu}\right)>1$. Hence, $|\lambda_1|>1$. We now demonstrate that $|\lambda_2|<1$. Similarly to the first case, define $\rho=-(\nu+\frac{1}{4}\frac{1}{\nu})$. Then,

$$|\lambda_2|<1 \qquad \text{if} \quad \left\{\rho-\sqrt{(\rho)^2-\frac{1}{2}}<1 \qquad \wedge \qquad -\rho+\sqrt{\rho^2-\frac{1}{2}}>-1\right\}.$$

This condition is always satisfied for $\rho > \frac{3}{4}$. However, since $\rho > 1$, this means that $|\lambda_1| < 1$ is always satisfied for $\nu < -1$. Then, also in this case, one eigenvalue of the system lies insider the unit circle and the other lies outside.

Hence, for any set of parameters satisfying Assumption S.3, the system the defines the sequential market equilibrium has one root inside the unit circle and one root outside. Hence, the equilibrium dynamics follow a saddle path that converges to the only steady state.

Proof of P.3. We now prove that the convergence path to the steady state follows cyclic dynamics. Specifically, $\forall t \in [0,\infty)$, i) if $\phi^Y_{C,t} < \tilde{\phi}^Y_C$, then $\phi^Y_{C,t+1} > \tilde{\phi}^Y_C$; and ii) if $\phi^Y_{C,t} > \tilde{\phi}^Y_C$, then $\phi^Y_{C,t+1} < \tilde{\phi}^Y_C$.

Assume $\phi^Y_{C,t-1} = \tilde{\phi}^Y_C + \kappa,\ \kappa \in \mathbb{R}$. By contraddiction, assume the statement is false in the sense that the convergence to the steady state is not cyclical. Then we would have: $\phi^Y_{C,t} = \tilde{\phi}^Y_C + \kappa' = \tilde{\phi}^Y_C + \delta_t \kappa$, for some $\delta_t \geq 0, \forall t$, so that $\kappa' < \kappa$. That is, the next-period point on the convergence path is always on the same "side" of the steady state. Then, the dynamic equation that describes the sequential market equilibrium can be written as:

$$\phi^Y_{C,t+1} - 2\alpha(\tilde{\phi}^Y_C + \delta_t \kappa) + \kappa(\tilde{\phi}^Y_C + \kappa) = \gamma.$$

Since $\tilde{\phi}^Y_C = \frac{\gamma}{1-2\alpha+\kappa}$, we have

$$\begin{aligned} \phi^Y_{C,t+1} &= \gamma \frac{1-2\alpha+\kappa+2\alpha-\kappa}{1-2\alpha+\kappa} + (2\alpha\delta_t - \kappa)\kappa \\ &= \tilde{\phi}^Y_C + (2\alpha\delta_t - \kappa)\kappa. \end{aligned} \tag{19}$$

Notice that the term $(2\alpha\delta_t - \kappa)$ is always negative, since $\alpha < 0,\ \kappa > 0,\ \delta_t \geq 0$. Thus, $\delta_{t+1} < 0$. As a result, the next-period point on the convergence path taken at t must be on the other "side" of the steady state, relative to the initial point $\tilde{\phi}^Y_C + \delta_t \kappa$. This results violates the initial assumption, which must thus be false. But since we only assumed that $\delta_t, \geq 0$, it must be that either $\delta_t < 0$, or $\delta_t > 0$ but not all following $\delta_t \geq 0$. Iterating the reasoning backwards by checking the case where we reinterpret the initial distance from the steady state as $\kappa = \tilde{\kappa}\delta_{t-1} = \tilde{\kappa}\frac{\kappa+\delta_t}{2\alpha}$. Hence, we can always find another initial condition such that δ_t must be of opposite sign relative to δ_{t-1}. Hence, $\delta_t < 0$ and $\delta_{t+1} > 0$, meaning that δ_t has alternating signs. That is: if $\phi^Y_{C,t} > \tilde{\phi}^Y_C$, then $\phi^Y_{C,t+1} < \tilde{\phi}^Y_C$. And if $\phi^Y_{C,t} < \tilde{\phi}^Y_C$, then $\phi^Y_{C,t+1} > \tilde{\phi}^Y_C$.

We now prove that the sequence $\left\{\left|\phi^Y_{C,t} - \tilde{\phi}^Y_C\right|\right\}_{t\geq 1}$ is monotonically decreasing and converging to zero. Suppose that the proposition is false. That is, there exists at least one $\tau > t$ such that $\left|\phi^Y_{C,\tau} - \tilde{\phi}^Y_C\right| - \left|\phi^Y_{C,\tau-1} - \tilde{\phi}^Y_C\right| > 0$. Then, from equation (19), it must be that $|\delta_\tau \kappa| > |\delta_{\tau-1}\kappa|$. That is,

$$|2\alpha\delta_{\tau-1} - \kappa| > |\delta_{\tau-1}|.$$

However, consider the implications at $\tau+1$. If $|\delta_\tau| > |\delta_{\tau-1}|$, then in $\tau+1$, it must also be the case that $|\delta_{\tau+1}| > |\delta_\tau|$. Hence, $\lim_{\tau\to\infty}(\delta_\tau) = \pm\infty$. This is impossible, since we have established that the system has a unique, saddle-path stable steady state. Hence, the initial statement must be false and the convergence to the steady state must be monotonic.

In the case with equality, we have that $2\alpha\delta_{\tau-1} - \kappa = \delta_{\tau-1}$. However, this would mean that $\lim_{t\to\infty}(\delta_t) = \pm\delta_{\tau-1}$, again inconsistent with the saddle-path stability of the system.

Hence, it must be $|\delta_{\tau+1}| < |\delta_\tau|\ \forall \tau \geq t, \forall t \geq 0$. Thus, the sequence $\{|\phi^Y_{C,t} - \tilde{\phi}^Y_C|\}_{t\geq 1}$ monotonically con-

verges to zero:

$$
\begin{aligned}
|\phi^Y_{C,t} - \tilde{\phi}^Y_C| &= |\tilde{\phi}^Y_C + \delta_t \kappa - \tilde{\phi}^Y_C| \\
&= |\delta_t \kappa| < |\delta_{t-1} \kappa| \\
&= |\phi^Y_{C,t-1} - \tilde{\phi}^Y_C|, \ \forall t \geq 1 \\
&\implies \begin{cases} \lim_{t \to \infty} |\phi^Y_{C,t} - \tilde{\phi}^Y_C| = 0. \\ |\phi^Y_{C,t} - \tilde{\phi}^Y_C| < |\phi^Y_{C,t-1} - \tilde{\phi}^Y_C|, \forall t \geq 1 \end{cases}
\end{aligned}
$$

□

B.6 Dynamics

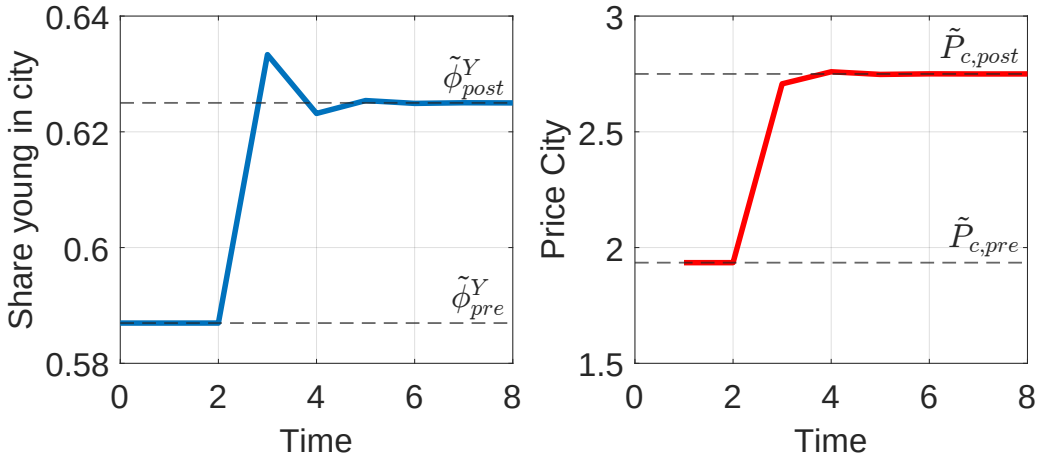
Proposition 3 indicates that the economy always converges to a steady state following a saddle-path dynamic. This convergence occurs cyclically, highlighting the presence of inter-generational spatial congestion. The central mechanism behind this phenomenon is that high-wage locations, which are desired by both generations, also foster high-skill development. When one "fortunate" generation accumulates more human capital, it is followed by an "unfortunate" generation crowded out of these high-wage, high-skill-generating locations by the older cohort, which is richer in skills, and as more incentives to locate in the city.

We illustrate these cyclical dynamics in Figure 15 and describe their connection to intergenerational congestion. In Panel (a), we depict the impact of an unanticipated permanent shock on the returns to experience in the city (e_C), which occurs in period 3. As a result, the young workers in period 3 have strong incentives to relocate to the city. Furthermore, since the previous generation of young workers in period 2 did not foresee the shock, the period 3 young workers face minimal competition for city space from the relatively low-skilled older generation. Consequently, spatial intergenerational congestion is low. This leads to the emergence of a"fortunate" generation in period 3, where more young people move to the city and acquire human capital at a higher rate compared to the new steady state $\tilde{\phi}^Y_{\text{post}}$. However, in period 4, this fortunate generation ages and becomes a highly skilled older generation. Their elevated levels of human capital motivate them to remain in the city, which in turn drives up housing prices. As a result, more young people find themselves unable to move to the city. Thus, every "fortunate" generation is eventually followed by an "unfortunate" generation, affected by intergenerational congestion in the housing market, which results in lower human capital accumulation. It is important to note that this unfortunate generation is still affected by the increased permanent shock to returns on experience. However, unlike the previous cohort, they must now compete for space with a wealthier, older generation.

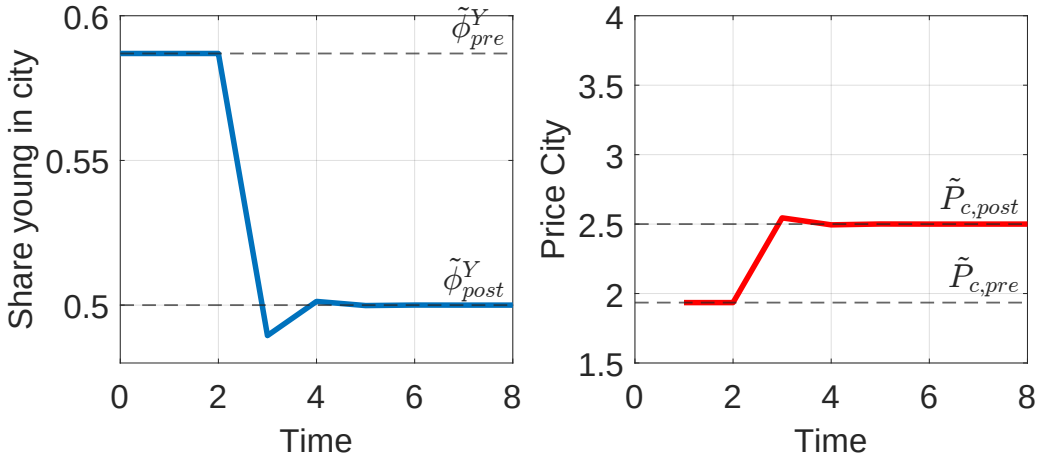
Panel (b) illustrates the effects of a permanent and unexpected increase in the unconditional return to experience, denoted as e. Similar to the previous example, this shock takes place in period 3. During this period, older workers unexpectedly find themselves equipped with higher skills, leading to an increased demand for the city. This surge in demand drives up house prices and reduces the number of young individuals who can afford to move to the city, pushing them below the new steady-state level due to intergenerational congestion. As a result, the young generation in period 3 will be relatively low-skilled when they age. Consequently, the young generation in period 4 experiences less intergenerational congestion, which enables them to accumulate higher skills.

Figure 15: Dynamics

(a) Shock to City-specific return to experience, e_C



(b) Shock to unconditional return to experience, e



Notes: Panel (a) shows the effect of a shock permanently increasing the return to the city e_C by 100 percentage points (from 1 to 2) in period 3 on the share of young (left plot) and the city price (right plot). Panel (b) shows the effect of a shock permanently increasing the return to global experience e from 0 to 1 in period 3. The other parameters are assumed as $\sigma = 2$, $Z_R = 1$, $\Delta Z = 1$. After the shock, the economy cycles around the new steady state share of young ($\tilde{\phi}_{\text{post}}$) and price ($\tilde{P}_{post}$) due to the presence of intergenerational congestion for city space.

B.7 Proof of Proposition 2

First, we prove two intermediate results that we will use throughtout the proof of the planner problem.

Lemma 4. *The average idiosyncratic utility of an individual i in age group g with skill s, given that they optimally choose their location j to maximise their utility, is*

$$\mathbb{E}(\varepsilon_{ij,t}|s_{it} = s, j = \arg\max V^g(s,\varepsilon_t)) = \sigma\phi_C^g(s) - \sigma(\phi_C^g(s))^2 + \mathbb{E}(\varepsilon_t|s,g)$$

Proof. Consider how $\varepsilon_C - \varepsilon_R \sim U(-\sigma,\sigma)$. Thus, we have that $\varepsilon_C - \varepsilon_R + \sigma \sim U(0,2\sigma)$. The average realised idiosyncratic utility of the individuals is given by the average realisation of the shock, which have common mean $\mathbb{E}(\varepsilon_t|s,g)$ as by assumption S.5, plus the integral - overall all individual with the

same skill and generation - of the individual realisations of $\varepsilon_C - \varepsilon_R$, for all individuals who chose the city rather than the rural location. Calling $\phi_C^g(s)$ the share of individuals of age g and skill s who choose $j = C$, we have that

$$\begin{aligned}\mathbb{E}(\varepsilon_{ij,t}|s_{it} = s, j = \arg\max V^g(s,\varepsilon_t)) &= \int_{2\sigma(1-\phi_C^g(s))}^{2\sigma} x\frac{1}{2\sigma}dx - \sigma\phi_C^g(s) + \mathbb{E}(\varepsilon_t|s,g) \\ &= \int_0^{2\sigma} x\frac{1}{2\sigma}dx - \int_0^{2\sigma(1-\phi_C^g(s))} x\frac{1}{2\sigma}dx - \sigma\phi_C^g(s) + \mathbb{E}(\varepsilon_t|s,g) \\ &= \sigma\phi_C^g(s) - \sigma(\phi_C^g(s))^2 + \mathbb{E}(\varepsilon_t|s,g),\end{aligned}$$

where the right-hand side equation in the first row comes from integrating $\varepsilon_C - \varepsilon_R + \sigma \sim U(0,2\sigma)$ over all individuals choosing the city, to which we need to substract the terms $\sigma\phi_C^g(s)$ so to be consistent with integrating over $\varepsilon_C - \varepsilon_R \sim U(-\sigma,\sigma)$. □

Lemma 5. ***Problem Equivalence.*** *Consider the identity ordering where workers of the same aged are clustered within the interval* $[0,2]$ *in each period, so that all the young always belong to the subinterval* $[0,1]$ *and all the old always belong to the subinterval* $[1,2]$*. Within each interval, individuals are ordered by their marginal contribution to welfare when relocated from the rural location to the city, that is their realisation of* $(\mu_C - \mu_R + \varepsilon_{i,C} - \varepsilon_{i,R} + (Z_C - Z_R)s_i)$*.*

Consider the alternative problem of a planner which chooses a cut-off ϕ^Y *such that:*

$$\hat{V}\left(\mathbb{S}, \{\varepsilon_i\}_{i\in[0,2]}\right) = \max_{\phi^Y\in[0,1]} \int U_i(s_i, j_i)di + \mathbb{E}\left[\hat{V}\left(\mathbb{S}', \{\varepsilon_i'\}_{i\in[0,2]}\right)\right] \quad s.t.$$

$$j_i = \begin{cases} C & if\ i > \phi^Y, i \in [0,1] \\ R & if\ i < \phi^Y, i \in [0,1] \\ C & if\ i-1 > 1-\phi^Y, i \in [1,2] \\ R & if\ i-1 < 1-\phi^Y, i \in [1,2] \end{cases}$$

$$\mathbb{S}' = \begin{cases} s^Y & with\ mass\ 1 \\ (1+e+e_C)s^Y & with\ mass \quad \phi^Y \\ (1+e+e_R)s^Y & with\ mass \quad (1-\phi^Y). \end{cases}$$

$$\phi_C^Y + \phi_C^O = 1.$$

Then, if

$$\exists\{\hat{c}_j^g(s)\}_{j\in\{C,R\},g\{Y,O\}} > 0 : \begin{cases} \phi_C^O(s) = \frac{1}{2} + \frac{1}{2\sigma}\left(\Delta\mu^O + c_C^O(s) - c_R^O(s)\right) \\ \phi_C^Y = \frac{1}{2} + \frac{1}{2\sigma}\left(\Delta\mu^Y + \Delta c^Y\right) + \frac{1}{2\sigma}\left[c_R^O(s^O(C)) - c_R^O(s^O(R)) \right. \\ \qquad \left. + \phi_C^O(s^O(C))\Delta c^O(s^O(C)) - \phi_C^O(s^O(R))\Delta c^O(s^O(R))\right] \end{cases},$$

the solution to this problem is equivalent to the planner problem's presented in Equation (2)*.*

Proof. Consider how the optimal location choice rule of an old worker implies that, for any two consumption bundles $\{c_C^O(s), c_R^O(s)\}$, a share $\phi_C^O(s)$ of old workers will choose the city, characterised as

$$\begin{aligned}\phi_C^O(s) &= \frac{1}{2} + \frac{1}{2\sigma}\left(\Delta\mu^O + c_C^O(s) - c_R^O(s)\right) \\ &= \frac{1}{2} + \frac{1}{2\sigma}\left(\Delta\mu^O + \Delta c^O(s)\right)\end{aligned}$$

Similarly, the share of young workers in the city, for any tuple of consumption bundles $\{c_j^g(s)\}_{j\in\{C,R\},g\{Y,O\}}$ is

$$\begin{aligned}\phi_C^Y =& \frac{1}{2}+\frac{1}{2\sigma}\left(\Delta\mu^Y+\Delta c^Y\right)\\ &+\frac{1}{2\sigma}\left[\left(\phi_C^O(s^O(C))c_C^O(s^O(C))+(1-\phi_C^O(s^O(C)))c_R^O(s^O(C))\right)\right.\\ &\qquad\left.-\left(\phi_C^O(s^O(R))c_C^O(s^O(R))+(1-\phi_C^O(s^O(R)))c_R^O(s^O(R))\right)\right]\\ =&\frac{1}{2}+\frac{1}{2\sigma}\left(\Delta\mu^Y+\Delta c^Y\right)\\ &+\frac{1}{2\sigma}\left[c_R^O(s^O(C))-c_R^O(s^O(R))+\phi_C^O(s^O(C))\Delta c^O(s^O(C))-\phi_C^O(s^O(R))\Delta c^O(s^O(R))\right]\end{aligned}$$

Notice that the individuals going in the city, for each generation and skill s, are the ones with the highest realisation of idiosyncratic preferences.

Moreover, notice that it must be optimal to set $\Delta c^O(C)-\Delta c^O(R)=\Delta Ze_C$, so that the marginal skilled and unskilled old that choose to go to the city provide the same marginal welfare contribution to the economy (in terms of both resources and idiosyncratic preferences). If this was not the case, at the margin the would be able to increase total welfare through a marginal decrease in either the skilled or unskilled city-rural consumption gap, compensated by an equivalent increase in the other skill group's consumption. Hence, it would not be a welfare-maximising consumption allocation.

This means any welfare-maximising consumption tuple implements:

1. A cut-off index for the young ϕ^Y, such that all young individuals indexed by $i<\phi^Y$ are allocated to the rural location, and all young individuals indexed by $i\geq\phi^Y$ are allocated to the city
2. A cut-off index for the old $1+\phi^O$, such that all old individuals indexed by $i<\phi^O$ are allocated to the rural location, and all old individuals indexed by $i\geq\phi^O$ are allocated to the city

That is, each individual i's location j_i is:

$$j_i=\begin{cases}C & \text{if } i>\phi^Y, i\in[0,1]\\ R & \text{if } i<\phi^Y, i\in[0,1]\\ C & \text{if } i-1>1-\phi^Y, i\in[1,2]\\ R & \text{if } i-1<1-\phi^Y, i\in[1,2]\end{cases}.$$

This cut-off must be welfare-maximising given the ordering of the agents in the $[0,2]$ interval determined by their realisation of $(\mu_C-\mu_R+\varepsilon_{i,C}-\varepsilon_{i,R}+(Z_C-Z_R)s_i)$. This statement can be easily proven. Ranking workers by $(\mu_C-\mu_R+\varepsilon_{i,C}-\varepsilon_{i,R}+(Z_C-Z_R)s_i)$ is equivalent to ranking them based on their overall contribution to welfare when assigned to location C instead or R, for the following reasons:

1. Since utility is linear in consumption, the increase in resources gained by relocating worker i from the rural area to the city, expressed as $(Z_C-Z_R)s_i$, is equivalent to the welfare increase that the planner can achieve by reallocating these resources for the consumption of any agent, minus the changes in the idiosyncratic and amenity components.
2. Old workers only care about the current period. Hence their contribution to welfare by being in the city is fully characterized by $(\mu_C-\mu_R+\varepsilon_{i,C}-\varepsilon_{i,R}+(Z_C-Z_R)s_i)$;

3. Young individuals have the same initial skill s^Y and face, conditional on their youth location j^Y, the same distributions of old-age returns to experience and idiosyncratic preferences. This means they all have the same expected future utility for a given level of consumption and realization of idiosyncratic utilities. Therefore, if we order the young individuals by the expression $(\mu_C - \mu_R + \varepsilon_{i,C} - \varepsilon_{i,R} + (Z_C - Z_R)s_i)$, we will achieve the same ranking as ordering them by their overall effect on the present value of welfare from being in the city.

Hence, directly choosing the welfare-maximising cutoff ϕ^Y and - thus - agents' locations according to their ordering in the $[0,2]$ interval determined by their realisation of $(\mu_C - \mu_R + \varepsilon_{i,C} - \varepsilon_{i,R} + (Z_C - Z_R)s_i)$ must yield the same maximum welfare as choosing the optimal consumption tuple, as long as there is a feasible, non-negative consumption tuple that implements such cutoffs when agents are free to choose their location. That is, the original planner's problem and the one where a planner directly choose agents' locations (via a single cut-off rule ϕ^Y for the young) under an incentive-compatibility constraint are identical. □

We now move on to proving Proposition 2.

Proof. *Proof* of P.7.
Recall the planner problem in Equation (2):

$$\begin{aligned}
\hat{V}(\mathbb{S}) = &\max_{\{c_j^g(s)\}_{j\in\{C,R\},g\{Y,O\}}} \int \mathbb{E}[U_i(s_i,j_i)]\,di + \mathbb{E}\left[\hat{V}\left(\mathbb{S}', \{\varepsilon_i'\}_{i\in[0,2]}\right)\right] \text{ s.t.} \\
n_j = &\int \mathbb{1}\{j_i = j\}di \quad j \in \{C,R\}, \\
\int Z_{j_i} s_i di = &\int_{i:g_i=Y} c_{j_i}^Y(s_i)di + \int_{i:g_i=O} c_{j_i}^O(s_i)di \\
\mathbb{S} \equiv &\{s_i\}_{i\in[0,2]}, \\
\mathbb{S}' = &\begin{cases} s^Y & \text{with mass 1} \\ (1+e+e_C)s^Y & \text{with mass } \int \mathbb{1}\{j_i = C\}\,di \\ (1+e+e_R)s^Y & \text{with mass } \int \mathbb{1}\{j_i = R\}\,di \end{cases}
\end{aligned}$$

Identity ordering. As a first step, let us cluster workers of the same age within the interval $[0,2]$ in each period, so that all the young always belong to the subinterval $[0,1]$ and all the old always belong to the subinterval $[1,2]$. Within each interval, we order individuals by their marginal contribution to welfare when relocated from the rural location to the city, that is their realisation of $(\mu_C - \mu_R + \varepsilon_{i,C} - \varepsilon_{i,R} + (Z_C - Z_R)s_i)$.

Using Lemma 5, we can recast the planner's problem in terms of selecting the optimal allocation *shares* of workers to the city. For ease of notation, let ϕ^O and ϕ^Y denote the steady-state share of old and young workers in the city, respectively. Then, the solution to the planner's problem is equivalent to the solution

to the following problem:

$$\hat{V}\left(\mathbb{S},\{\varepsilon_i\}_{i\in[0,2]}\right)=\max_{\phi^Y\in[0,1]}\int U_i(s_i,j_i)di+\mathbb{E}\left[\hat{V}\left(\mathbb{S}',\{\varepsilon_i'\}_{i\in[0,2]}\right)\right]\quad\text{s.t.}$$

$$j_i=\begin{cases}C & \text{if } i>\phi^Y, i\in[0,1]\\ R & \text{if } i<\phi^Y, i\in[0,1]\\ C & \text{if } i-1>1-\phi^Y, i\in[1,2]\\ R & \text{if } i-1<1-\phi^Y, i\in[1,2]\end{cases}$$

$$\mathbb{S}'=\begin{cases}s^Y & \text{with mass } 1\\ (1+e+e_C)s^Y & \text{with mass } \phi^Y\\ (1+e+e_R)s^Y & \text{with mass } (1-\phi^Y).\end{cases}$$

$$\phi_C^Y+\phi_C^O=1$$

$$\exists\{\hat{c}_j^g(s)\}_{j\in\{C,R\},g\{Y,O\}}>0:\begin{cases}\phi_C^O(s)=\frac{1}{2}+\frac{1}{2\sigma}\left(\Delta\mu^O+c_C^O(s)-c_R^O(s)\right)\\ \phi_C^Y=\frac{1}{2}+\frac{1}{2\sigma}\left(\Delta\mu^Y+\Delta c^Y\right)+\frac{1}{2\sigma}\left[c_R^O(s^O(C))-c_R^O(s^O(R))\right.\\ \qquad\left.+\phi_C^O(s^O(C))\Delta c^O(s^O(C))-\phi_C^O(s^O(R))\Delta c^O(s^O(R))\right]\end{cases}$$

We will proceed as follows: First, we will state the first-order condition (FOC). Then, we will derive it and show that it is both necessary and sufficient for identifying the maximum. Finally, we will analytically derive the unique solution and prove that there exists a set of feasible and incentive-compatible consumption tuples that yield such allocation even when the planner is not allowed to directly relocate individuals.

Taking the FOC:

$$\frac{\partial}{\partial\phi^Y}\int U_i(s_i,j_i)di+\frac{\partial}{\partial\phi^Y}\mathbb{E}\left[\hat{V}\left(\mathbb{S}',\{\varepsilon_i'\}_{i\in[0,2]}\right)\right]=0. \tag{20}$$

This proof aims to derive each component of the first-order condition and express it as a function of ϕ_Y.

First component: present utility of old workers. At any given time, the planner considers the utility of the young and old alive in that period. Let us decompose the integral of the present period utility into the ones for each location/age pair, i.e.:

$$\begin{aligned}\int U_i(s_i,j_i)di=&\left(\int_{i:j_i=C}U_i^Y(s^Y,C)di+\int_{i:j_i=R}U_i^Y(s^Y,R)di\right)\\&+\left(\int_{i:j_i=C}U_i^O(s_i,j_i)di+\int_{i:j_i=R}U_i^O(s_i,j_i)di\right).\end{aligned} \tag{21}$$

Let us start from the old workers' component. We can rewrite:

$$\left(\int_{i:j_i=C}U_i^O(s_i,j_i)di+\int_{i:j_i=R}U_i^O(s_i,j_i)di\right)=\int_{2-\phi^O}^{2}U_i^O(s_i,C)di+\int_1^{2-\phi^O}U_i^O(s_i,R)di. \tag{22}$$

Let denote with θ^O the share of old workers with high skill, at a given period.[36] For the moment we take it as given, and later we will link it to the planner's optimal spatial allocation of young workers. Also, recall that $\phi^O(s^O(C))$ denotes the share of skilled old in the city, and $\phi^O(s^O(R))$ denotes the share of

[36] That is with skill $1+e+e_C$.

unskilled old in the city. By definition, the total share of old allocated to the city, ϕ^O, is:

$$\phi^O \equiv \phi^O(s^O(C))\theta^O + \phi^O(s^O(R))(1-\theta^O).$$

Now, recall that workers are ranked by their marginal contribution to welfare when allocated to the city instead of the rural location. In terms of resources, an unskilled older worker allocated to the city always contributes less to welfare, quantified as $\Delta Z e_C$, compared to a skilled older worker with the same idiosyncratic preferences. According to Assumption S.2, the difference in idiosyncratic components is distributed as $\Delta\varepsilon \sim U(-\sigma,\sigma)$. Therefore, at the optimum, given a probability $\phi^O(s^O(C))$ that a skilled individual is assigned to the city, the probability that an unskilled older individual is allocated to the city must be $\frac{\Delta Z e_C}{2\sigma}$ smaller.[37]

Hence,

$$\phi^O(s^O(C)) = \phi^O(s^O(R)) + \frac{\Delta Z e_C}{2\sigma}. \tag{23}$$

This result implies the following relationships between the optimal spatial allocation $\phi^O(\cdot)$ of workers with both skill levels and their shares, θ^O and $1-\theta^O$, taking the latter as given (for the moment):

$$\begin{aligned}
\phi^O(s^O(R)) &= \phi^O - \theta^O \Delta Z \frac{e_C}{2\sigma} \\
\phi^O(s^O(C)) &= \frac{1}{\theta^O}\left[\phi^O - \phi^O((1+e))(1-\theta^O)\right] \\
&= \phi^O + (1-\theta^O)\Delta Z \frac{e_C}{2\sigma}.
\end{aligned} \tag{24}$$

Notice that equation (24) implies that $d\phi^O(s^O(\cdot))/d\phi^O = 1$.

With this information, we can decompose the marginal effect of increasing the share of old workers in the city on welfare into the marginal effect of bringing more high-skilled and low-skilled workers into the city.

Specifically, we can break down each integral in equation (22) into two components: one that considers only the skilled old, represented by a mass of θ^O, and another that considers only on the unskilled old, with a mass of $1-\theta^O$. It is also important to note that each worker's utility in a location is independent of the proportion of workers assigned to that location; that is, $dU_i^O(s_i,j)/d\phi^O = 0$.[38] Thus, by applying the Leibniz rule to equation (22), we obtain that the change in utility from increasing the share of old workers in the city is represented by the weighted average—between the skilled and unskilled groups—of

[37] When ordering all old workers by their welfare contribution, the unskilled workers with the highest realization of idiosyncratic utility for the city, which is $\Delta\varepsilon = \sigma$ from Assumption S.2, are ranked lower than any skilled old worker with realization $\Delta\varepsilon \geq \sigma - \Delta Z e_C$. This means that a share $\frac{\Delta Z e_C}{2\sigma}$ of high-skilled workers is always allocated to the city before any unskilled workers. Hence, for any share of old ϕ^O, the share of high-skilled individuals allocated to the city is $\phi^O(s^O(C)) = \phi^O(s^O(R)) + \frac{\Delta Z e_C}{2\sigma}$.

[38] Since the planner allocates the output to the workers, and individual output does not depend on the population size or age composition of the location, changing ϕ^O has no effect on $U_i^O(s_i,j)$.

the change in utility that occurs when we move the marginal old worker from region R to city C. That is:

$$\begin{aligned}
&\frac{\partial}{\partial \phi^O}\left(\int_{2-\phi^O}^{2} U_i^O(s_i,C)di+\int_{1}^{2-\phi^O} U_i^O(s_i,R)di\right)= \\
&\theta^O\left(\Delta\mu^O+\Delta Z(1+e+e_C)-\underbrace{\left(-\sigma+2\sigma\phi^O(s^O(C))\right)}_{\text{Diff. in idiosyncratic utility of the skilled}}\right) \\
&+(1-\theta^O)\left(\Delta\mu^O+\Delta Z(1+e)-\underbrace{\left(-\sigma+2\sigma\phi^O(s^O(R))\right)}_{\text{Diff. in idiosyncratic utility of the unskilled}}\right), \\
=&\Delta\mu^O+\Delta Z(1+e)-\sigma\left(-1+2\phi^O(s^O(C))\theta^O+2\phi^O(s^O(R))(1-\theta^O)\right)+\theta^O\Delta Ze_C, \\
=&\Delta\mu^O+\Delta Z(1+e+e_C)-\sigma\left(-1+2\phi^O\right)-(1-\theta^O)\Delta Ze_C.
\end{aligned}$$

Because of the OLG structure of the model, the share of the skilled olf worker in the current period, θ^O is equal to the share of young workers allocated in the city in the previous period, which at the steady state is always costant and equal to ϕ^Y. Using the fact that $\phi^O=1-\phi^Y$, we have:

$$\begin{aligned}
&\frac{\partial}{\partial \phi^O}\left(\int_{2-\phi^O}^{2} U_i^O(s_i,C)di+\int_{1}^{2-\phi^O} U_i^O(s_i,R)di\right) \\
&\quad=\Delta\mu^O+\Delta Z(1+e+e_C)-\sigma\left(1-2\phi^Y\right)-(1-\phi^Y)\Delta Ze_C.
\end{aligned}$$

Since $\partial\phi^O/\partial\phi^Y=-1$,

$$\begin{aligned}
&\frac{\partial}{\partial \phi^Y}\left(\int_{2-\phi^O}^{2} U_i^O(s_i,C)di+\int_{1}^{2-\phi^O} U_i^O(s_i,R)di\right)= \\
&\quad-\left(\Delta\mu^O+\Delta Z(1+e+e_C)-\sigma\left(1-2\phi^Y\right)-(1-\phi^Y)\Delta Ze_C\right),
\end{aligned} \tag{25}$$

which is an expression only on ϕ^Y.

Second component: present utility of young workers. Going back to equation (21), and following the same reasoning, the derivative of the first-period utility of the young is:

$$\frac{\partial}{\partial \phi^Y}\left(\int_{i:j_i=C} U_i^Y(s^Y,C)di+\int_{i:j_i=R} U_i^Y(s^Y,R)di\right)=\Delta\mu^Y+\Delta Z+\sigma\left(1-2\phi^Y\right), \tag{26}$$

which is also an expression only on ϕ^Y.

Third component: expected future value of the planner. Finally, we need to derive the expectation term of the first-order condition, $\frac{\partial}{\partial\phi^Y}\mathbb{E}\left[\hat{V}\left(\mathbb{S}',\{\varepsilon_i'\}_{i\in[0,2]}\right)\right]$.

Today's share of the young in the city, ϕ^Y, affects future expected utility only through the overlapping generation structure of the model, as the young today will become old next period. Because of our human accumulation assumption, the current share of young in the city, ϕ^Y, is identical to the share of of highly skilled old next period, which we denoted $\theta^{O'}$. Hence,

$$\frac{\partial}{\partial \phi^Y}\mathbb{E}\left[\hat{V}\left(\mathbb{S}',\{\varepsilon_i'\}_{i\in[0,2]}\right)\right]=\frac{\partial}{\partial \theta^{O'}}\mathbb{E}\left[\hat{V}\left(\mathbb{S}',\{\varepsilon_i'\}_{i\in[0,2]}\right)\right].$$

Since the problem is recursive:

$$\hat{V}\left(\mathbb{S}', \{\varepsilon_i'\}_{i\in[0,2]}\right) = \max_{\phi^{\prime Y}\in[0,1]} \int U_i(s_i', j_i')di + \mathbb{E}\left[\hat{V}\left(\mathbb{S}'', \{\varepsilon_i''\}_{i\in[0,2]}\right)\right] \implies$$
$$\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\hat{V}\left(\mathbb{S}', \{\varepsilon_i'\}_{i\in[0,2]}\right)\right] = \frac{\partial}{\partial \theta^{O\prime}} \max_{\phi^{\prime Y}\in[0,1]} \mathbb{E}\left\{\int U_i(s_i', j_i')di + \mathbb{E}\left[\hat{V}\left(\mathbb{S}'', \{\varepsilon_i''\}_{i\in[0,2]}\right)\right] | \phi^Y\right\}.$$

The following-period expectation, $\mathbb{E}\left[\hat{V}\left(\mathbb{S}'', \{\varepsilon_i''\}_{i\in[0,2]}\right)\right]$, does not depend on $\theta^{O\prime}$, since the skill level of the old in a period has no effect on the next-period skill distribution. Hence, $\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\hat{V}\left(\mathbb{S}'', \{\varepsilon_i''\}_{i\in[0,2]}\right)\right] = 0$. Decomposing $\int U_i(s_i', j_i')di$ into the next-period istantaneous utility of old and young, we have:

$$\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\int U_i(s_i', j_i')di\right] = \frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\int_0^1 U_i^Y(s_i', j_i')di\right] + \frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\int_1^2 U_i^O(s_i', j_i')di\right].$$

The next period instantaneous utility of the young, $U_i^Y(\cdot)$, depends on their consumption, amenities, and idiosyncratic utility realization. Since each worker's utility is linear in consumption and the planner assigns equal welfare weight to all agents, the planner is indifferent between any allocation that meets the resource constraints. For simplicity, we assume that the planner distributes to each agent a level of consumption equal to their production. Moreover, both amenities and the idiosyncratic utility realization depend only on the current spatial allocation, not on the share of skilled old workers in the economy. Hence, $\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\int_0^1 U_i^Y(s_i', j_i')di\right] = 0$. Thus, the next-period derivative only depends on how a change in the next-period skill distribution of the old affect their overall next-period instantaneous utility through their consumption (that is, the resources they generate), amenities, and idiosyncratic utility:

$$\begin{aligned}
\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left[\hat{V}\left(\mathbb{S}', \{\varepsilon_i'\}_{i\in[0,2]}\right)\right] &= \frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\int_{i\in[1,2]} U_i^O(s_i', j_i')di] \\
&= \frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\underbrace{\left(\int_{i\in[1,2]} Z_{j_i'} s_i' di\right)}_{\text{Resources}} + \frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\underbrace{\left(\int_{i\in[1,2]} \mu_{j_i'}^O di\right)}_{\text{Amenities}} \\
&+ \frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\underbrace{\left(\int_{i\in[1,2]} \varepsilon_{i,j_i'}' di\right)}_{\text{Idiosyncratic Utility}}.
\end{aligned}$$

We derive each of these terms.

Amenities. Since the amenities are the same across all skill levels, $\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left(\int_{i\in[1,2]} \mu_{j_i'}^O di\right) = 0$.

Resources. We now focus on the derivative of tomorrow's expected resources, $\frac{\partial}{\partial \theta^{O\prime}}\mathbb{E}\left(\int_{i\in[1,2]} Z_{j_i'} s_i' di\right)$. Moving the marginal young from rural to city today increases the share of next-period skilled old work-

ers, with an expected change in *future* resources equal to:

$$\begin{aligned}\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int_{i\in[1,2]}Z_{j_i'}s_i'di|\phi^Y\right)=&\frac{\partial}{\partial\theta^{O'}}\left[\mathbb{E}\int_{s_i'=s^O(C)}Z_{j_i'}s^O(C)di+\mathbb{E}\int_{s_i'=s^O(R)}Z_{j_i'}s^O(R)di\right]\\=&\frac{\partial}{\partial\theta^{O'}}\left[\theta^{O'}\mathbb{E}\left(Z_{j_i'}s_i'|s_i'=s^O(C)\right)+(1-\theta^{O'})\mathbb{E}\left(Z_{j_i'}s_i'|s_i'=s^O(R)\right)\right]\\=&\mathbb{E}(Z_{j_i'}|s^O(C))s^O(C)-\mathbb{E}(Z_{j_i'}|s^O(R))s^O(R)+\\&+\theta^{O'}\frac{\partial\mathbb{E}\left(Z_{j_i'}|s^O(C)\right)s^O(C)}{\partial\theta^{O'}}+(1-\theta^{O'})\frac{\partial\mathbb{E}\left(Z_{j_i'}|s^O(R)\right)s^O(R)}{\partial\theta^{O'}}.\end{aligned}\tag{27}$$

Using equation (24), the expected productivity of a worker in the next period is:

$$\begin{aligned}\mathbb{E}(Z_{j_i'}|s^O(C))=\phi^{O'}(s^O(C))Z_R+(1-\phi^{O'}(s^O(C)))Z_C=Z_R+\Delta Z\left[\phi^{O'}+(1-\theta^{O'})\Delta Z\frac{e_C}{2\sigma}\right],\\\mathbb{E}(Z_{j_i'}|s^O(R))=\phi^{O'}(s^O(R))Z_R+(1-\phi^{O'}(s^O(R)))Z_C=Z_R+\Delta Z\left[\phi^{O'}-\theta^{O'}\Delta Z\frac{e_C}{2\sigma}\right],\end{aligned}$$

with their derivatives being:

$$\frac{\partial}{\partial\theta^{O'}}\mathbb{E}(Z_{j_i'}|s^O(C))=\frac{\partial}{\partial\theta^{O'}}\mathbb{E}(Z_{j_i'}|s^O(R))=-\frac{\Delta Z^2e_C}{2\sigma}.$$

Plugging these expressions into equation (27) yields the marginal gain in resources obtained by moving the marginal young to the city:

$$\begin{aligned}\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int_{i\in[1,2]}Z_{j_i'}s_i'di|\phi^Y\right)=&Z_Re_C+(\Delta Z)^2\frac{e_C}{2\sigma}(1+e)+\phi^{O'}\Delta Ze_C+(1-\theta^{O'})\frac{(\Delta Z)^2(e_C)^2}{2\sigma}+\\&-\theta^{O'}\frac{(\Delta Z)^2(e_C)^2}{2\sigma}-\frac{\Delta Z^2e_C}{2\sigma}(1+e)\\=&e_C\left(Z_R+\phi^{O'}\Delta Z+(1-2\theta^{O'})\frac{(\Delta Z)^2e_C}{2\sigma}\right)\\=&e_C\left(Z_R+\phi^{O'}\Delta Z+(1-2\phi^{Y})\frac{(\Delta Z)^2e_C}{2\sigma}\right),\end{aligned}\tag{28}$$

where the last equivalence comes from the fact that $\theta^{O'}=\phi^Y$ because of the OLG structure of the model.

Idiosyncratic Utility. Finally, we look at the marginal effect of moving more young to the city on tomorrow's realized idiosyncratic utility: $\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int_{i\in[1,2]}\varepsilon'_{i,j_i'}di\right)$. Following the same steps as for Equation (27), the change in the average realization of the idiosyncratic utility component is:

$$\begin{aligned}\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int\limits_{i\in[1,2]}\varepsilon'_{i,j_i'}di\right)=&\frac{\partial}{\partial\theta^{O'}}\left[\theta^{O'}\mathbb{E}(\varepsilon'_{i,j_i'}|s_i'=s^O(C))+(1-\theta^{O'})\mathbb{E}(\varepsilon'_{i,j_i'}|s_i'=s^O(R))\right]\\=&\mathbb{E}(\varepsilon'_{i,j_i'}|s_i'=s^O(C))-\mathbb{E}(\varepsilon'_{i,j_i'}|s_i'=s^O(R))+\\&+\theta^{O'}\frac{\partial\mathbb{E}(\varepsilon'_{i,j_i'}|s_i'=s^O(C))}{\partial\theta^{O'}}+(1-\theta^{O'})\frac{\partial\mathbb{E}(\varepsilon'_{i,j_i'}|s_i'=s^O(R))}{\partial\theta^{O'}}.\end{aligned}\tag{29}$$

Recall that, as shown in Lemma 4, a given skill level s, the conditional expectations of the realisation of

the idiosyncratic shocks is:

$$\mathbb{E}(\varepsilon'_{i,j'_i}|s'_i=s)=\underbrace{\sigma\phi^{O'}(s)-\sigma(\phi^{O'}(s))^2}_{\mathbb{E}\left(\varepsilon'_{i,j'_i}-\varepsilon'_{i,\tilde{j}'_i}|s'_i=s\right)}+\mathbb{E}(\varepsilon'|s'_i=s).$$

Thus, the difference between the expectations of skilled and unskilled old is

$$\mathbb{E}(\varepsilon'_{i,j'_i}|s^O(C))-\mathbb{E}(\varepsilon'_{i,j'_i}|s^O(R))=\frac{\Delta Ze_C}{2}\left(1-\frac{\Delta Ze_C}{2\sigma}-2\phi^{O'}+\frac{\Delta Ze_C}{\sigma}\theta^{O'}\right)+\Xi, \tag{30}$$

where we have used that $\phi^{O'}(s^O(C))=\phi^{O'}(s^O(R))+\frac{\Delta Ze_C}{2\sigma}$ from equation (23) and that $\phi^{O'}(s^O(R))=\phi^{O'}-\theta^{O'}\frac{\Delta Ze_C}{2\sigma}$, from equation (24) and $\Xi=\mathbb{E}(\varepsilon'|s'_i=s^O(C))-\mathbb{E}(\varepsilon'|s'_i=s^O(R))$.

Then, using Liebniz rule we derive

$$\begin{aligned}
&\theta^{O'}\frac{\partial\mathbb{E}(\varepsilon'_{i,j'_i}|s'_i=s^O(C))}{\partial\theta^{O'}}+(1-\theta^{O'})\frac{\partial\mathbb{E}(\varepsilon'_{i,j'_i}|s'_i=s^O(R))}{\partial\theta^{O'}}\\
&\quad=\theta^{O'}\frac{\partial\left(\sigma\phi^{O'}(C)-\sigma(\phi^{O'}(C))^2+\mathbb{E}(\varepsilon'|s^O(C))\right)}{\partial\phi^{O'}(C)}\frac{\partial\phi^{O'}(C)}{\partial\theta^{O'}}\\
&\qquad+(1-\theta^{O'})\frac{\partial\left(\sigma\phi^{O'}(R)-\sigma(\phi^{O'}(R))^2+\mathbb{E}(\varepsilon'|s^O(R))\right)}{\partial\phi^{O'}(R)}\frac{\partial\phi^{O'}(R)}{\partial\theta^{O'}}\\
&\quad=-\frac{\Delta Ze_c}{2}+\Delta Ze_C\phi^{O'},
\end{aligned}$$

where we have used that $\phi^{O'}(s^O(C))=\phi^{O'}(s^O(R))+\frac{\Delta Ze_C}{2\sigma}$ from equation (23) and that $\phi^{O'}(s^O(R))=\phi^{O'}-\theta^{O'}\frac{\Delta Ze_C}{2\sigma}$, from equation (24).

Plugging equations (30) and (**??**) into equation (29), we have:

$$\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int_{i\in[1,2]}\varepsilon'_{i,j'_i}di\right)=-\frac{\Delta Ze_c}{2}+\Delta Ze_C\phi^{O'}+\frac{\Delta Ze_C}{2}\left(1-\frac{\Delta Ze_C}{2\sigma}-2\phi^{O'}+\frac{\Delta Ze_C}{\sigma}\theta^{O'}\right)+\Xi.$$

Because of the OLG structure of the model, $\theta^{O'}=\phi^Y$. Also, in the steady state, $\phi^{O'}=\phi^O=\frac{1-\phi^Y}{\alpha}$. Therefore,

$$\begin{aligned}
\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int_{i\in[1,2]}\varepsilon'_{i,j'_i}di\right)=&-\frac{\Delta Ze_c}{2}+\Delta Ze_C\frac{1-\phi^Y}{\sigma}\\
&+\frac{\Delta Ze_C}{2}\left(1-\frac{\Delta Ze_C}{2\sigma}-2\frac{1-\phi^Y}{\sigma}+\frac{\Delta Ze_C}{\sigma}\phi^Y\right)+\Xi.
\end{aligned} \tag{31}$$

Finally, we derive Ξ. Intuitively, when the planner switches a young from living in the countryside to the city, they not only enhance their skills but also modify the expected future utility. This utility change can be expressed as by $\Xi=(\mathbb{E}(\varepsilon_j-\varepsilon_{\tilde{j}}|s_i=s^O(R))-\mathbb{E}(\varepsilon_j-\varepsilon_{\tilde{j}}|s_i=s^O(C)))$. To make the planner problem

comparable to the market economy in terms of this shift in utility, we compute Ξ as the utility offset in the steady-state market equilibrium, i.e.

$$\begin{aligned}\Xi &= \left(\mathbb{E}(\Delta\varepsilon'_{i,j'_i}|s_i = s^O(R)) - \mathbb{E}(\Delta\varepsilon'_{i,j'_i}|s_i = s^O(C))\right), \\ &= \left(\sigma\tilde{\phi}^{O'}(R) - \sigma(\tilde{\phi}^{O'}(R))^2\right) - \left(\sigma\tilde{\phi}^{O'}(C)) - (\sigma\tilde{\phi}^{O'}(C))^2\right).\end{aligned}$$

It is convenient to rewrite the right hand side as a function of $\tilde{\phi}^Y$ which is the market steady state allocation of the young to the city:

$$\Xi = -\frac{\Delta Ze_C}{2}\left(1 - \frac{\Delta Ze_C}{2\sigma} - 2(1-\tilde{\phi}^Y) + \frac{\Delta Ze_C}{\sigma}\tilde{\phi}^Y\right).$$

Substituting this expression into equation (31):

$$\frac{\partial}{\partial\theta^{O'}}\mathbb{E}\left(\int_{i\in[1,2]}\varepsilon'_{i,j'_i}di\right) = \left(\phi^Y - \tilde{\phi}^Y\right)\left(\Delta Ze_C + \frac{\Delta Z^2e_C^2}{2\sigma}\right) + \frac{1}{2}\Delta Ze_C - \Delta Ze_C\phi^Y \tag{32}$$

Notice that we were able to write this component as a function of ϕ^Y, but also on the market steady state share $\tilde{\phi^Y}$. Recall that this dependence depends solely on our preference assumptions, which allow for clean and transparent comparison between the market and the planner allocations.

Summing all components: We have proved that, at the steady state, the planner's FOC in equation (20) is equivalent to setting the sum of the right hand side of (25), (26), (28) and (32) equal to zero, which is:

$$\begin{aligned}&\underbrace{(\Delta\mu^Y - \Delta\mu^O) - \Delta Ze - \Delta Ze_C\phi^Y - 2\sigma\left(2\phi^Y - 1\right)}_{\text{Change in current utility}} + \underbrace{\left(\phi^Y - \tilde{\phi}^Y\right)\left(\Delta Ze_C + \frac{\Delta Z^2e_C^2}{2\sigma}\right) + \frac{1}{2}\Delta Ze_c - \phi^Y\Delta Ze_C}_{\text{P.V. of future idiosyncratic utility change}} \\ &+ \underbrace{e_C\left(Z_R + (1-\phi^Y)\Delta Z + (1-2\phi^Y)\frac{(\Delta Z)^2(e_C)}{2\sigma}\right)}_{\text{P.V. of future resource change}} = 0.\end{aligned} \tag{33}$$

Notice that the FOC depends only on ϕ^Y.

Second order condition From equation (33), we derive the second order condition of the steady state planner problem. Taking the derivative with respect to ϕ^Y, and accounting for how the planner cannot change the initial condition, the second order condition is:

$$\frac{\partial^2\hat{V}\left(\mathbb{S}, \{\varepsilon_i\}_{i\in[0,2]}\right)}{(\partial\phi^Y)^2} = 2\Delta Ze_C - 8\sigma < 0,$$

where the inequality comes from the fact that $\sigma > |\Delta Ze_C|$, as by Assumption S.3. Hence, the one-shot welfare function is concave in ϕ^Y and the first order condition of the planner problem is both sufficient and necessary.[39]

[39] An intuition for why the one-shot welfare function is concave is that, since the idiosyncratic preference shock follows a continuous and monotonically increasing distribution, the set of individuals with the same marginal

Solution. Solving for $\hat{\phi}^Y$, we obtain the planner's optimal share of young allocated to the city:

$$\begin{aligned}\hat{\phi}^Y =& \frac{1}{2} + \frac{(\Delta\mu^Y - \Delta\mu^O)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}} + \frac{Z_R e_C - \Delta Z e}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}} \\ & - \frac{\left(\tilde{\phi}^Y - \frac{1}{2}\right)\left(\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}\right)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}}.\end{aligned}$$

Calling $\tilde{\phi}^Y_{\psi=0}$ the market allocation of young to the city in the case $\psi = 0$, the planner's solution can be written as:

$$\begin{aligned}\hat{\phi}^Y =& \tilde{\phi}^Y_{\psi=0} + \underbrace{\frac{\left(\tilde{\phi}^Y_{\psi=0} - \tilde{\phi}^Y\right)\left(\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}\right)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}}}_{\approx 0} \\ \approx & \tilde{\phi}^Y_{\psi=0}.\end{aligned}$$

Necessary conditions. As a penultimate step, we prove that $(\Delta Z e_C > 0)$ is a sufficient and necessary condition for $\hat{\phi}^Y > \tilde{\phi}^Y$.

Proof of P.6. The total output of the economy depends on the labour market parameters (Z_R, ΔZ, e_C, e) and the equilibrium allocation of young and old (skilled and unskilled) to the city:

$$\begin{aligned}\Pi =& \phi^Y(Z_R + \Delta Z) + (1-\phi^Y)Z_R + \phi^Y\left(\phi^O(C)(Z_R + \Delta Z) + (1-\phi^O(C))Z_R\right)(1 + e + e_C) \\ & + (1-\phi^Y)\left(\phi^O(R)(Z_R + \Delta Z) + (1-\phi^O(R))Z_R\right)(1+e) \\ =& Z_R + \Delta Z + \phi^Y Z_R e_C + Z_R(1+e) + \phi^Y \Delta Z(e + e_C)\frac{\Delta Z e_C}{2\sigma} + \phi^Y\phi^O(R)\Delta Z e_C + \phi^O(R)\Delta Z e.\end{aligned}$$

Taking the difference between the output under the planner's solution ($\hat{\imath}$) and under the market's steady state ($\tilde{\imath}$), we obtain

$$\begin{aligned}\hat{\imath} - \tilde{\imath} =& (\hat{\phi}^Y - \tilde{\phi}^Y)\left(Z_R e_C + \Delta Z(e + e_C)\frac{\Delta Z e_C}{2\sigma} - \Delta Z e\left(1 + \frac{\Delta Z e_C}{2\sigma}\right)\right) + \left(\hat{\phi}^Y\hat{\phi}^O(R) - \tilde{\phi}^Y\tilde{\phi}^O(R)\right)\Delta Z e_C \\ =& (\hat{\phi}^Y - \tilde{\phi}^Y)\left(Z_R e_C + \Delta Z(e + e_C)\frac{\Delta Z e_C}{2\sigma} - \Delta Z e\left(1 + \frac{\Delta Z e_C}{2\sigma}\right) + \Delta Z e_C - \Delta Z e_C\left(1 + \frac{\Delta Z e_C}{2\sigma}\right)(\hat{\phi} + \tilde{\phi})\right) \\ =& (\hat{\phi}^Y - \tilde{\phi}^Y)\left(Z_R e_C - \Delta Z e_C + \Delta Z e_C\left(1 + \frac{\Delta Z e_C}{2\sigma}\right)(1 - \hat{\phi} - \tilde{\phi})\right).\end{aligned}$$

Notice that $\hat{\phi} + \tilde{\phi} < 2$, as they represent share of young in the city with value in the interior of $(0,1)$. Hence, $(1 - \hat{\phi} - \tilde{\phi}) > -1$. Moreover, recall that $\psi > 0 \implies \hat{\phi} > \tilde{\phi}$, as proved above. Hence, $\psi > 0$ and $Z_R e_C > \Delta Z\left(e + e_C\left(1 + \frac{\Delta Z e_C}{2\sigma}\right)\right)$ is a sufficient condition for $\hat{\imath} - \tilde{\imath} > 0$, for any value of the amenities parameters.

contribution to aggregate welfare has mass zero. Hence, the gain from moving the marginal young into the city, at the cost of moving out the marginal old, is strictly decreasing in $\hat{\phi}^Y$: all youngs have the same skill, and decreasing levels of net idiosyncratic utility for the city. Conversely, the welfare loss from moving the marginal old individual out of the city is always increasing, as we would be moving out individuals with increasing levels of idiosyncratic utility and - potentially - increasing skill levels.

□

B.8 Pareto Improvement

We have shown how the market allocation is potentially inefficient, relative to the one of the planner. We now show that, starting from the market steady state, there exists a Pareto-improving allocation which implements the planner's allocation as long as the young's income is sufficiently front-loaded and large relative to the amenity value of the city.

We also make a final assumption to provide a clearer insight into the nature of the misallocation. In particular, we provide an alternative assumption to Assumption S.4: we assume that the difference in the mean idiosyncratic shocks between the high-skilled and low-skilled types are always equal to the offset necessary to satisfy Assumption S.4 when $\psi = 0$, or equivalently $\psi^Y = \psi^O = 1$.

Assumption 2.

S.6 *(Centering - Alternative) Call $\tilde{\phi}^Y_{\psi=0}$ the steady-state share of young in the city in the equilibrium where $\psi^Y = \psi^O = 0$. Then, for any skill, $\varepsilon_{i,j}$ and $\varepsilon_{i,\bar{j}}$ satisfy:*

$$\mathbb{E}_t(\varepsilon_{t+1}|s^O(j^Y) = s^O(C)) - \mathbb{E}_t(\varepsilon_{t+1}|s^O(j^Y) = s^O(R)) = -\frac{\Delta Z e_C}{2}\left(1 - \frac{\Delta Z e_C}{2\sigma} - 2(1-\tilde{\phi}^Y_{\psi=0}) + \frac{\Delta Z e_C}{\sigma}\tilde{\phi}^Y_{\psi=0}\right).$$

Corollary 6. *Assume Assumptions S.1, S.2, S.3, S.5, and S.6 hold. Also, assume that $n_C = n_R = 1$, $P_R = 1$, and $s^Y = 1$.*

P.7 *Assume Assumption S.6 holds. Call $\tilde{\phi}^Y_{\psi=0}$ the steady-state share of young in the city in the equilibrium where $\psi^Y = \psi^O = 0$. Then, for any value of ψ,*

$$\hat{\phi}^Y = \tilde{\phi}^Y_{\psi=0}$$

Proof. Consider how, under assumption S.6, the offset to be applied to workers' idiosyncratic utilities, Ξ, is now the same regardless of the current share of young in the city, ϕ^Y_t, as we take as reference point the hypothetical market steady state where $\psi = 0$. That is, the offset does not change any more as a function of the ϕ^Y realised in equilibrium. Otherwise, the planner's problem is the same. Hence, the FOC of the planner's problem is

$$\underbrace{(\Delta\mu^Y - \Delta\mu^O) - \Delta Z e - \Delta Z e_C \phi^Y - 2\sigma\left(2\phi^Y - 1\right)}_{\text{Change in current utility}} + \underbrace{\left(\phi^Y - \tilde{\phi}^Y_{\psi=0}\right)\left(\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}\right) + \frac{1}{2}\Delta Z e_c - \phi^Y \Delta Z e_C}_{\text{P.V. of future idiosyncratic utility change}}$$

$$+ \underbrace{e_C\left(Z_R + (1-\phi^Y)\Delta Z + (1-2\phi^Y)\frac{(\Delta Z)^2(e_C)}{2\sigma}\right)}_{\text{P.V. of future resource change}} = 0.$$

This yields the optimal share of young in the city

$$\begin{aligned}\hat{\phi}^Y =& \frac{1}{2} + \frac{(\Delta\mu^Y - \Delta\mu^O)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}} + \frac{Z_R e_C - \Delta Z e}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}} \\ & - \frac{\left(\tilde{\phi}^Y_{\psi=0} - \frac{1}{2}\right)\left(\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}\right)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}}. \\ =& \tilde{\phi}^Y_{\psi=0} + \underbrace{\frac{\left(\tilde{\phi}^Y_{\psi=0} - \tilde{\phi}^Y_{\psi=0}\right)\left(\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}\right)}{4\sigma + 2\Delta Z e_C + \frac{\Delta Z^2 e_C^2}{2\sigma}}}_{=0} \\ =& \tilde{\phi}^Y_{\psi=0}.\end{aligned}$$

□

This result shows that the planner's allocation is the one replicating the allocation of the market when $\psi = 0$. That is, when the marginal cost of space is identical for the young and the old. The intuition behind this result is that, when $\psi \neq 0$, old and young individuals face different costs from the presence of intergenerational congestion. Mathematically, in this economy the same good - housing - has two prices, one for the young and one for the old, not necessarily aligned to the individual or social returns. Due to the difference in price, an increase in demand for the city discourages less old than young to move away, changing the city composition. However, nothing ensures that the young leaving the city have a lower lifetime consumer surplus from living in the city than the old who remain, since they face different prices. Since the market does not sort according to the individual contribution to total welfare, but rather the individual consumption net of housing costs, a planner can always do better than the market unless $\psi = 0$.

We can now prove the following result.

Proposition 7. *Assume the economy is in steady-state, so that the initial share of skilled old is $\theta^O = \tilde{\phi}^Y$ and the steady state price of the city is $\tilde{P}$. Call $\hat{P}$ the steady-state price of an economy in steady state where $\psi^O = \psi^Y = 1$. Call $\hat{s} = (1+e+e_C)$ if $\psi > 0$, or $\hat{s} = (1+e)$ if $\psi < 0$. Call $\{\hat{c}^g_{j,t}(s)\}_{g\in\{Y,O\},j\in\{C,R\},t\geq 1}$ the planner's optimal consumption bundle. Define τ^g as*

$$\tau^g = \begin{cases} (\Delta\mu^O + \Delta Z\hat{s} + \tilde{P}_C - \hat{P}_C)(\hat{\phi}^Y - \tilde{\phi}^Y)\left(1 + \frac{\Delta Z e_C}{2\sigma}\right) + \frac{1}{2}\tilde{P}_C\left((1-\psi)(1-\tilde{\phi}^Y) + \tilde{\phi}^Y\right) - \frac{1}{2}\tilde{P}_C \\ \qquad + \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi} + \tilde{\phi})\right) \\ \qquad + (\hat{\phi}^Y - \tilde{\phi}^Y)(1 - \frac{\Delta Z e_C}{2\sigma} + 2\frac{\Delta Z e_C}{2\sigma}(\hat{\phi}^Y - \tilde{\phi}^Y) + \frac{\Delta Z^2 e_C^2}{4\sigma^2}(\tilde{\phi}^Y + \hat{\phi}^Y)) & \text{if } g{=}O \\ -\tau^O & \text{if } g{=}Y. \end{cases}$$

Then, if

$$\begin{cases} \min\{\hat{c}^Y_C, \hat{c}^Y_R\} \geq -\tau^Y, \\ \min\{\hat{c}^O_C(s), \hat{c}^O_R(s)\} \geq -\tau^O, & \forall s \\ \hat{c}^O_R \geq (\hat{\phi} - \tilde{\phi}^Y)\Delta Z e_C, & \text{if } \psi > 0, \\ \hat{c}^O_C \geq -(\hat{\phi} - \tilde{\phi}^Y)\Delta Z e_C, & \text{if } \psi < 0, \end{cases}$$

there exists a feasible Pareto-improving allocation $\{c^g_{j,t}(s)\}_{g\in\{Y,O\},j\in\{C,R\},t\geq 1}$ that:

1. *Implements the planner's optimum share of young in the city, $\hat{\phi}^Y$, in every period $t \geq 1$.*

2. *The consumption allocation satisfies, for each individual in generation g, with skill s, in location j at time t,*

$$
\begin{aligned}
c^g_{j,t}(s) &= \hat{c}^g_j(s) + \tau^g, \qquad \forall s, \forall (t,g) : (t \geq 2, g \in \{Y,O\}) \wedge (t=1, g=Y),\\
c^O_{C,1}(s) &= \hat{c}^O_C(s) + \tau^O + (1-\hat{\phi}^O(s))(\hat{\phi}-\tilde{\phi}^Y)\Delta Z e_C\\
&\quad - \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi}+\tilde{\phi})\right),\\
c^O_{R,1}(s) &= \hat{c}^O_R(s) + \tau^O - \hat{\phi}^O(s)(\hat{\phi}-\tilde{\phi}^Y)\Delta Z e_C\\
&\quad - \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi}+\tilde{\phi})\right).
\end{aligned}
$$

3. *For each individual in generation g, with skill s, at time t - before the idiosyncratic shocks ε_{ijt} are realised - the ex-ante expected utility of every agent is higher or equal than under the market allocation, with every generation besides the initial old one ($g = O$ at $t = 1$) being strictly better off:*

$$
\begin{aligned}
\mathbb{E}_1\left(V^O(s,\varepsilon_t)|\{c^g_{j,t}(s)\}\right) &\geq \mathbb{E}_1\left(V^O(s,\varepsilon_t)|\{\tilde{c}^g_{j,t}(s)\}\right),\ \forall s, t=1\\
\mathbb{E}_t\left(V^Y(s,\varepsilon_t)|\{c^g_{j,t}(s)\}\right) &> \mathbb{E}_t\left(V^Y(s,\varepsilon_t)|\{\tilde{c}^g_{j,t}(s)\}\right),\ \forall s, \forall t.
\end{aligned}
$$

Proof. Consider the expected steady-state utility of an old individual from the first generation, of skill s. This can be expressed as

$$
\begin{aligned}
\mathbb{E}(\tilde{V}^O_1(s)) &= \tilde{\phi}^O(s)\left(\mu^O_C + \tilde{c}^O_C(s)\right) + (1-\tilde{\phi}^O(s))(\mu^O_R + \tilde{c}^O_R(s))\\
&= \tilde{\phi}^O(s)\left[\mu^O_C + (Z_R+\Delta Z)s + \tilde{\tau} - \psi^O\tilde{P}\right] + (1-\tilde{\phi}^O(s))\left[\mu^O_R + Z_R s + \tilde{\tau}\right]\\
&= \mu^O_R + \tilde{\phi}^O(s)\Delta\mu^O + Z_R s + \tilde{\phi}^O(s)\Delta Z s + \tilde{\tau} - \tilde{\phi}^O(s)\psi^O\tilde{P}.
\end{aligned}
$$

Then, we guess and verify that the allocation of consumption $\{c^g_{j,t}(s)\}_{g\in\{Y,O\},j\in\{C,R\},t\geq 1}$ such that

$$
\begin{aligned}
c^g_{j,t}(s) &= \hat{c}^g_j(s) + \tau^g, \qquad \forall s, \forall (t,g) : (t \geq 2, g \in \{Y,O\}) \wedge (t=1, g=Y),\\
c^O_{C,1}(s) &= \hat{c}^O_C(s) + \tau^O + (1-\hat{\phi}^O(s))(\hat{\phi}-\tilde{\phi}^Y)\Delta Z e_C\\
&\quad - \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi}+\tilde{\phi})\right),\\
c^O_{R,1}(s) &= \hat{c}^O_R(s) + \tau^O - \hat{\phi}^O(s)(\hat{\phi}-\tilde{\phi}^Y)\Delta Z e_C\\
&\quad - \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi}+\tilde{\phi})\right).
\end{aligned}
\tag{34}
$$

is a feasible Pareto improvement whenever $\min\{Z_C - \tilde{P}_{\psi=0}, Z_R\} \geq -\tau^Y_t$ and $Z_r(1+e) \geq (\hat{\phi}^Y - \tilde{\phi}^Y)\Delta Z e_C$.

First, we prove feasibility of city space. Consider how the allocation implements $\phi^Y_t = \hat{\phi}^Y, \forall t$. That is, the share of young in the city is always equal to the one of the old. However, the initial conditition is the market's steady state. Hence, the share of skilled old is $\theta^O = \tilde{\phi}^Y$. For the allocation to be feasible in terms of city space, it needs to satisfy

$$
\phi^O_{C,t}(s^O(C))\theta^O + \phi^O_{C,t}(s^O(R))(1-\theta^O) = 1 - \tilde{\phi}^Y.
$$

Substituting $\theta^O = \tilde{\phi}^Y$, and imposing the same gap between $\phi^O_{C,t}(s^O(C))$ and $\phi^O_{C,t}(s^O(R))$ as the market,

so that $\phi^O_{C,t}(s^O(C)) - \phi^O_{C,t}(s^O(R)) = \frac{\Delta Z e_c}{2\sigma}$, the condition implies

$$\phi^O_{C,t}(s^O(R)) = 1 - \hat{\phi}^Y - \tilde{\phi}^Y \frac{\Delta Z e_C}{2\sigma},$$
$$\phi^O_{C,t}(s^O(C)) = 1 - \hat{\phi}^Y + (1 - \tilde{\phi}^Y) \frac{\Delta Z e_C}{2\sigma}.$$

Call $\hat{\phi}^O_1(s)$ the allocation share of the old of skill s that satisfies Equation B.8. Notice that the planner's share of old of skill s in the city satisfies

$$\hat{\phi}^O_{C,t}(s^O(R)) = 1 - \hat{\phi}^Y - \hat{\phi}^Y \frac{\Delta Z e_C}{2\sigma},$$
$$\hat{\phi}^O_{C,t}(s^O(C)) = 1 - \hat{\phi}^Y + (1 - \hat{\phi}^Y) \frac{\Delta Z e_C}{2\sigma}.$$

Hence,

$$\begin{aligned}\hat{\phi}^O_{C,t}(s) - \hat{\phi}^O_1(s) &= (\tilde{\phi}^Y - \hat{\phi}^Y) \frac{\Delta Z e_C}{2\sigma} \\ &= \frac{1}{2\sigma} \left[(\hat{c}^O_C(s) - \hat{c}^O_R(s)) - (c^O_{C,1}(s) - c^O_{R,1}(s)) \right].\end{aligned}$$

This implies that any space-feasible allocation must implement a consumption gap between the city and rural location, for any first-period old individual with skill s, of

$$c^O_{C,1}(s) - c^O_{R,1}(s) = (\hat{c}^O_C(s) - \hat{c}^O_R(s)) - (\tilde{\phi}^Y - \hat{\phi}^Y) \Delta Z e_C.$$

Among all the allocations that satisfy this relationship, we look for the one that leaves the expected utility of individuals of skill s unchanged. That is, the one that satisfies for some scalar $x \in (0,1)$ the relationship

$$-\phi^O_1(s) x \left(\hat{\phi}^Y - \tilde{\phi}^Y \right) \Delta Z e_C + (1 - \phi^O_1(s))(1 - x) \left(\hat{\phi}^Y - \tilde{\phi}^Y \right) \Delta Z e_C = 0.$$

Solving for x, we find that $x = 1 - \phi^O_1(s)$ is the coefficient for which the consumption of space is feasible.

Second, we verify if the allocation is feasible in terms of consumption. Note that the gap in total production between the planner and the proposed allocation is caused by the fact that the share of old skilled is different between the two allocations, due to the different starting condition ($\hat{\phi}$ for the planner, $\tilde{\phi}$ for the economy starting at the market steady state). Taking the difference between the total production of the planner and the one of the proposed allocation, we get the following expression for the gap in resources between the two economies:

$$\begin{aligned}&Z_R e_C (\hat{\phi}^Y - \tilde{\phi}^Y) + \Delta Z e_C \frac{\Delta Z e_C}{2\sigma} (\tilde{\phi}^Y - \hat{\phi}^Y) + \Delta Z (1 + e + e_C) \left(\hat{\phi}^Y \hat{\phi}^O(C) - \tilde{\phi}^Y \phi^O(C) \right) \\ &+ \Delta Z (1 + e) \left(\tilde{\phi}^Y \phi^O(R) - \hat{\phi}^Y \hat{\phi}^O(R) \right). \\ &= \left(\tilde{\phi}^Y - \hat{\phi}^Y \right) \left[Z_R e_C + \frac{\Delta Z e_C}{2\sigma} \Delta Z e_C + \Delta Z e_C (1 - \hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma} \left(\hat{\phi}^Y + \tilde{\phi}^Y \right) \right]. \end{aligned} \tag{35}$$

Hence, aggregate consumption in the first generation must be equal to the consumption in the planner's economy, minus Equation (35). Summing the consumption allocations from Equation (34) for $t = 1$

indeed yields

$$\begin{aligned}\int c_{i,t=1}di = &\int \hat{c}^{g_i}_{j_i}(s_i)di \\ &\underbrace{+\left[\tilde{\phi}^Y\tau_t^O + (1-\tilde{\phi}^Y)\tau_t^O\right]}_{\text{Sum of }\tau_1^O} - \tau_t^Y \\ &-(\tilde{\phi}^Y + (1-\tilde{\phi}^Y))\left(\tilde{\phi}^Y - \hat{\phi}^Y\right)\left[Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}\left(\hat{\phi}^Y + \tilde{\phi}^Y\right)\right] \\ = &\int \hat{c}^{g_i}_{j_i}(s_i)di - \left(\tilde{\phi}^Y - \hat{\phi}^Y\right)\left[Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}\left(\hat{\phi}^Y + \tilde{\phi}^Y\right)\right].\end{aligned}$$

Hence, the proposed consumption allocation is feasible, in terms of aggregate consumption.

Finally, we need to check that consumption is non-negative for every single individuals. Given the qualifying assumptions of this proposition, it is clear that consumption is always non-negative within the assumed parameter space.

Hence, the proposed allocation is feasible.

Now, we prove that the allocation is a Pareto-improvement. Consider how, from $t \geq 2$, the allocation is exactly that of the planner. Since the planner is welfare-maximising, the total consumption - and thus welfare - achieved in each period must be strictly larger than that of the market for any $\psi \neq 0$, as long as the two locations are not exactly identical to each other. Since, in expectation, the first-period young receive when old the same transfer they performed when young, they are also better off than in the market allocation.

Hence, to conclude the proof, we only need to check whether the first generation old are not worse-off under the new allocation. Consider how, under the new allocation, their expected utility is the same as the one realised under the planner's allocation, besides the change in idiosyncratic utility given by the fact that $\hat{\phi}^O(s) \neq \phi^O(s)$. Hence, the expected value of the old in this allocation is

$$\begin{aligned}\mathbb{E}(V^O(\cdot)) = &\mathbb{E}(\hat{V}^O(\cdot)) + (\tilde{\phi}^Y - \hat{\phi}^Y)(1 - \frac{\Delta Z e_C}{2\sigma} + 2\frac{\Delta Z e_C}{2\sigma}(\hat{\phi}^Y - \tilde{\phi}^Y) + \frac{\Delta Z^2 e_C^2}{4\sigma^2}(\tilde{\phi}^Y + \hat{\phi}^Y)) \\ &-\left(\tilde{\phi}^Y - \hat{\phi}^Y\right)\left[Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1-\hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}\left(\hat{\phi}^Y + \tilde{\phi}^Y\right)\right] + \tau^O(s)\end{aligned}$$

The difference between this and the expected utility under the market allocation is

$$\begin{aligned}\mathbb{E}(V^O) - \mathbb{E}(\tilde{V}^O) = &\mathbb{E}(\hat{V}^O(\cdot)) - \mathbb{E}(\tilde{V}^O(\cdot)) + (\Delta\mu^O + \Delta Z s + \tilde{P}_C - \hat{P}_C)(\hat{\phi}^Y - \tilde{\phi}^Y)\left(1 + \frac{\Delta Z e_C}{2\sigma}\right) \\ &+ \frac{1}{2}\tilde{P}_C\left((1-\psi)(1-\tilde{\phi}^Y) + \tilde{\phi}^Y\right) - \frac{1}{2}\tilde{P}_C + \Delta Z(1 + e + e_C - s)(\hat{\phi} - \tilde{\phi}) \\ = &-\left[(\Delta\mu^O + \Delta Z s + \tilde{P}_C - \hat{P}_C)(\hat{\phi}^Y - \tilde{\phi}^Y)\left(1 + \frac{\Delta Z e_C}{2\sigma}\right) + \frac{1}{2}\tilde{P}_C\left((1-\psi)(1-\tilde{\phi}^Y) + \tilde{\phi}^Y\right) - \frac{1}{2}\tilde{P}_C\right] \\ &+ (\Delta\mu^O + \Delta Z s + \tilde{P}_C - \hat{P}_C)(\hat{\phi}^Y - \tilde{\phi}^Y)\left(1 + \frac{\Delta Z e_C}{2\sigma}\right) + \frac{1}{2}\tilde{P}_C\left((1-\psi)(1-\tilde{\phi}^Y) + \tilde{\phi}^Y\right) - \frac{1}{2}\tilde{P}_C \\ &+ \Delta Z(\hat{s} - s)(\hat{\phi} - \tilde{\phi}) \\ \geq &0,\end{aligned}$$

where the last inequality comes from the fact that if $\psi > 0$, then $\hat{\phi} > \tilde{\phi}$ and $\hat{s} = 1 + e + e_C \geq 1 + e$ imply that the last term is positive. And if $\psi < 0$, then then $\hat{\phi} < \tilde{\phi}$ and $\hat{s} = 1 + e \leq 1 + e + e_C$ imply that the last term is positive.

Hence, the first generation of skilled old is exactly as well off as under the market allocation, while the first generation of unskilled old is at least as well off. Hence, the proposed allocation is a Pareto-improvement. □

Proposition 7 states that, whenever the consumption of the young and in the rural location is high enough, it is possible to implement the planner's allocation from $t+1$, and the planner's share of young in the city, $\hat{\phi}^Y$, starting from the first period t. To do so, it is necessary to compensate the first-generation of old individuals (young individuals, for $\psi < 0$), who are worse-off relative to the market allocation. This compensation accounts for: i) the lower consumption when in the city (due to paying "full price", rather than $\psi^O P_C$), and iii) the lower probability of choosing the city due to the change in relative consumption (which reduces expected utility), which also affects the average realised idiosyncratic utility. In order to make the young better off, any reduction in consumtpion (relative to the planner's consumption allocation) is compensated for when old.[40]

In other words, as long as the income profile of the young is sufficiently front-loaded, a planner can allocate more young to the city, and compensate the old for their loss in utility relative to the market allocation. This allows to bring the economy to the planner's allocation after a single period, which has an overall higher level of welfare, relative to the market steady state, whenever $\psi \neq 1$. This Pareto-improving allocation can also be achieved by a policymaker who can perform age-location-skill-specific lump-sum transfers, as shown in the following proposition:

Proposition 8. *Consider a policymaker who can implement age-place-skill-specific transfers and taxes. Call $T^g_{jt}(s)$ the (net) transfer received by an individual with skill s, in location j, and of age g living in location j. Suppose the economy starts in steady state. Then, the following set of transfers implements the Pareto-improving allocation described in Proposition 7:*

$$
\begin{aligned}
T^g_{jt}(s) &= \left(\hat{c}^g_{jt}(s) - \tilde{c}^g_{jt}(s)\right) + \tau^g, \forall (t,g) : (t \geq 2, g \in \{Y,O\}) \wedge (t = 1, g = Y), \\
T^O_{C,1}(s) &= \left(\hat{c}^O_C(s) - \tilde{c}^O_C(s)\right) + \tau^O + (1 - \hat{\phi}^O(s))(\hat{\phi} - \tilde{\phi}^Y)\Delta Z e_C \\
&\quad - \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1 - \hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi} + \tilde{\phi})\right), \\
T^O_{R,1}(s) &= \left(\hat{c}^O_R(s) - \tilde{c}^O_R(s)\right) + \tau^O - \hat{\phi}^O(s)(\hat{\phi} - \tilde{\phi}^Y)\Delta Z e_C \\
&\quad - \left(\hat{\phi}^Y - \tilde{\phi}^Y\right)\left(Z_R e_C + \frac{\Delta Z e_C}{2\sigma}\Delta Z e_C + \Delta Z e_C(1 - \hat{\phi}^Y) - \frac{\Delta Z e_C}{2\sigma}(\hat{\phi} + \tilde{\phi})\right).
\end{aligned}
$$

That is, the economy sits at the planner's optimum at any $t \geq 2$, while no individual is worse off relative to the market's steady state.

[40] Notice how, while here we present *one* possible Pareto-improving allocation, many other allocations exist, with different within-generation transfer differences between high- and low-skilled and, thus, different outcomes in terms of within-generation inequality.

C Estimation

C.1 Borrowing and Interest Rates

We identify the parameter related to interest rates and borrowing directly from the data.

We impose a maximum collateral constraint of $M_p = 0.90$, equivalent to 90% LTV, as these represent the highest LTV ratio that has been consistently offered over the years in the UK.

We choose the income multiplier constraints to be $M_p = 3.5$, meaning that the maximum borrowing is 3.5 times the individual income, a common condition included in mortgage offers.

From the data, we fix q to be the average Bank of England policy rate for the period 2002-2019, which is equal to 2.07%.

We model the interest premium over borrowings ($b < 0$) as

$$f\left(\frac{-b}{p_{jt}h}\right) = q_1 + q_2 \min\left\{\max\left\{\frac{-b}{p_{jt}h} - l_m, 0\right\}, 1\right\}$$

Here, q_1 represents the spread between the risk-free rate q and the rate applied to a mortgage with LTV ratio $\frac{-b}{p_{jt}h} \leq l_m$, while q_2 represents the interest rate premium for a fully unsecured loan (LTV of 100%). Hence, mortgage with LTV$= M_p$ commands a total interest rate $q + q_1 + q_2(M_p - l_m)$. Consistently with the literature and data, we choose $l_m = 0.6$ as the LTV ratio below which further reducing leverage produces no interest rate reduction.

For the period 2012-2019, we find that the policy rate was 0.5%, the average quoted mortgage rate for 60% LTV was 1.843%, and the average 90% LTV rate was 3.416%. Hence,

$$\begin{aligned} q_1 &= 0.01843 - 0.005 = 0.01343, \\ q_2 &= \frac{0.03416 - q_1 - q_0}{M_p - l_m} = 0.0524. \end{aligned}$$

That is, the maximum interest rate premium for a fully-unsecured loan (100% LTV), relative to a 60% LTV loan, was approximately 2.2 percentage points. The premium for a 90% LTV loan was 1.6 pp.

C.2 Housing Depreciation

From ONS's National Account statistics on fixed capital assets and consumption, we estimate a depreciation rate of dwellings of 1.97%, not including land. The land underlying residential buildings is estimated to be worth £5,400bn by ONS, against the total value of dwellings being £4,407bn. Hence, the pure value of dwellings represents 44.9% of the joint value of dwellings and land. We assume, as in Greaney et al. (2025), that land does not depreciate. Thus, we set $\delta = 0.0197 \times 44.9\% = 0.895\%$ per year.

C.3 Estate agents costs

Rents are determined by the profit-maximising behaviour of real estate agents, which charge an optimal rental price

$$r_{jt} = (\delta + q_0 + q_1 + m)p_{jt} - \dot{p}_{jt}.$$

Hence, the estate agent costs m represent the extra cost of rental units relative to an equivalent real estate investment (with cost $(\delta + q_0 + q_1)p_{jt}$), net of price growth and depreciation. Importantly, since we have

assumed that the real estate sector is fully competitive, the calibrated level of m will also capture the tentants' costs associated with a lack of competitiveness in the sector.

Aggregating across locations and discretising, we have the discrete-time approximation for average national rents

$$\bar{r}_t \approx (\delta + q + m)\bar{p}_t - \Delta\bar{p}_{t+1}.$$

The parameters δ and $q = T^{-1}\sum_t q_t$ are known. In Section C.2 we estimate $\delta = 0.00895$, per year. For q, we take the average of the 2002-2019 period, equal to $2.07\% + 1.343\% = 3.413\%$.[41] We estimate average rents $\bar{r}_t$ using the English Housing Survey. Average prices $\bar{p}_t$ and average price growth $\Delta\bar{p}_{t+1}$ are estimated from the Land Registry price paid microdata, representing the universe of geo-coded housing transactions from 2004 to 2020, meaning that we have data on the 1-year forward variations for all property transactions between 2004 and 2019.

Specifically, we obtain average monthly rents for England from the historical chain-linked series of the Price Index of Private Rents and Private Housing Rental prices. We then use the index's 2015 regional, property type (flats, terraced, semi-detached or detached) and furnished status weights to calculate the share of unfurnished and furnished properties used to calculate the average rental price. We find that furnished properties are given a 77% weight. From anecdotical evidence in the literature and industry knowledge, we consider furnished properties to be rented out at a 20% Furnished Premium. Hence, we calculate an unfurnished-normalised rental index as

$$r_{t,\text{norm}} = \frac{r_t}{\text{Unfurnished Share} + (1 + \text{Furnished Premium}) \times (1 - \text{Unfurnished Share})}.$$

Then, we construct an index of property prices. The main issue we need to address is that the distribution of transaction prices across regions and types may not be the same as the one of rental properties. Hence, we calculate property prices from 2005 to 2020 by region and type using transaction data from the Price Paid dataset provided by the Land Registry. Then, using the microdata on individual transactions, we construct a new (unfurnished) rent-consistent property price index for each month, weighting each transaction so that their distribution by region and type is consistent with the Private Rent Index weights. This yields a property price level normalised for comparison with the unfurnished-normalised rental price data.

Finally, we approximate property price changes $\Delta\bar{p}_{t+1}$ as the *real* price change between the same month of the following year:

$$\Delta\bar{p}_{t+1} = \left(\frac{p_{t+1}}{\text{CPI}_{t+1}} - \frac{p_t}{\text{CPI}_t}\right)\text{CPI}_t.$$

Finally, using the time series for rents, prices, and price changes derived above, we can recover m by inverting equation C.3:

$$m = \frac{1}{T}\sum_{t\in T}\left(\frac{\bar{r}_t + \Delta\bar{p}_{t+1}}{\bar{p}_t}\right) - \delta - q_0 - q_1 \approx 0.0174.$$

We include in our sample all months from January 2005 to December 2022. Our estimate puts the overall costs of the rental company at $m = 1.74\%$ of the value of the property. To account for the fact that we do

[41] Notice that, in our estimating equation, the different q_t average out to q. Hence, we do not need to know its time series.

not include land rent for leasehold properties and community expenses in our model, usually paid by the landlord and thus accounted for in the rent calculation, we subtract them from the estimates for m.

Land rents. In England and Wales, most apartment buildings and many recent houses are built under a property structure known as "leasehold". In a leasehold, owners buy the right to freely use the *interior* of the building, but do not own the land, the walls, roof, nor - in the case of apartament blocks - the shared parts such as corridors and staircases. Leaseholders thus pay a yearly "ground rent" to the landowner.[42] Around 20% of dwellings in the UK are leaseholds and, in 2022, the average ground rent was around 0.1% of the value of the property. Hence, to account for ground rents, we reduce m by the average ground rent calculated over all England and Wales property, or 0.02 pp. of their average value.

Service charges. In addition to ground rents, leasehold properties and condominiums pay community expenses to a management company to maintain the common areas and services, commonly denoted as "service charge". The average service charge in 2025 was £2300 per leasehold flat, equivalent to £1696 in 2015 terms, or about 0.7% of the property value. Accounting for the average share of leasehold properties on the market, this corresponds to a further downward adjustment of 0.14 pp. for m.

While leaseholds are somewhat more prevalent in the rental sector, we do not make any further adjustment to correctly capture the difference in costs of the average rental dwelling to the one of the average owner-occupied dwelling. Hence, our final estimate for m corresponds to $m = 0.0158$.

That is, for a £300,000 property, the additional rental costs over that of a homeowner with a fully secured mortgage is approximately £4740 a year. While somewhat large relative to the US literature (Greaney et al., 2025), notice that - probably not incidentally - this spread is close to the estimated interest rate premium of a fully unsecured mortgage, relative to a secured one. In fact, the premium of a 90% LTV loan vs a 60% LTV loan is $q_2(M_p - l_m) = 0.0162$. This could be an indication that the rental sector is actually not fully competitive, with landlords maximising rents by setting them to the level that would prevent low-wealth individuals from purchasing a property, or that the marginal landlord is a leveraged loan-to-let mortgager.[43]

C.4 Council Tax

England's Council Tax (and its commercial property equivalent, the Business Rates) is a tax levied on the *occupiers* of residential properties. Landlords only pay the tax if the property is vacant. The structure of the tax is set by national directives, with tax "bands" assigned according to the 1991 assessment of the value of a property. Local councils can choose the level of the central band of the tax ("band D"), but not the principles on which to apply exemptions and scaling with the value of the property. All properties built after 1991 are taxed according to their 1991-based theoretical price, calculated based on neighbouring properties' prices and hedonic adjustments.

Although its criterions are set nationally, the revenue of the tax goes to local authorities to finance public services. Hence, the tax is not proportional to the property value. The tax is capped at two times the tax applied to the central band for properties valued more than £320k in 1991, while the minimum band is half (below £40k properties in 1991 prices) of the central band. In terms of share of property value, the

[42] Ground rents have been abolished in 2022 for new dwellings only. Notice that our model is calibrated to 2015 data.

[43] Indeed, in the UK several small owners have purchased rental properties with "loan-to-let" schemes, often on interest-only mortgages, which they aim to repay with the proceedings of rental costs, which - together with the expected real housing price growth - are in expectation at least as large as the opportunity cost of alternative investments and the cost of a leverage mortgage. Once again, justifying $m \approx q_2(M_p - l_m)$.

average tax is around 0.68% of the property value for a property just qualifying for the lowest band but only 0.27% for a property just qualifying for the highest band.

Since 2016 house prices were approximately three times 1991 house prices, we calculate the minimum council tax as $0.0068 \times 40,000 \times 3/27,615 = 0.0295$ income units in the model, or £816 in 2015 terms. We set the variable component to $(0.0027 \times 960,000 - 816)/960,000 = 0.00185$, or a 0.185% proportional tax on the property's value.

C.5 Housing size

We then use data from the Household Cost Index to obtain estimate of household expenditure on housing costs, including rents, social rents, bills, interest rates on mortgages, owner-occupied recurrent housing costs (e.g. service charges), and maintenance. We exclude the stamp duty, the tax paid upon buying a property, as in our model it is represented by other costs. We find that the average cost of housing is approximately 23.5% of housing expenditure, and hence set σ, the parameter capturing the share of non-housing consumption, to $\sigma = 1 - 0.235 = 0.765$.

We pin down rent per unit by normalising the average price of a property such that $h = 1$ corresponds to a £50,000 property in the average location. Hence, at an average price of £2750/sqm obtained from Land Registry data for 2015, it corresponds to £50,000/(£2750/sqm)=18.2 sqm with a normalised price $\overline{p} = 50,000/27615 = 1.81$.[44] Since in steady-state

$$r_j = (q_0 + q_1 + \delta + m + \tau_{r0})p_j + \tau_{r1},$$

we recover the average rent consistent with our normalisation of $h = 1$ as $\overline{r} = 0.111$.

From the Land Registry, we calculate that the 1st percentile of houses sold in the UK is 36.0sqm, equivalent to $h = 1.97$. The 95th percentile of house sizes sold every year is 174sqm, equivalent to $h = 9.6$. Since most buyers of such large properties are couples, we normalise $H_{\max} = 5.0$, approximately equal to an *individual* maximum house value of £250,000 in the average-price location (which, for example, would correspond to nearly £500,000, or a £1m house bought as a couple, in the London TTWA).

C.6 Distance Matrix

We calibrate the distance matrix across locations by calculating the travel time between population centroids at peak hour using Google Maps. We normalise the maximum distance to 1, and own-distance (from location j to itself) to zero.

Then, call T_j the set of TTWAs that constitute the model location j, and $T_{j'}$ the set of TTWAs that constitute the model location j'. To find the distance between j and j' we take the weighted average between all bilateral distances between each element of the sets T_j and $T_{j'}$, where the weights are given by the product of the origin and destination TTWA population, normalised to that their sum across all T_j, $T_{j'}$ element pairs is equal to 1. Finally, we set the own-distance of j, $D_{jj'}$, equal to zero and normalise the largest distance observed in the distance matrix to 1.

C.7 Learning Function

C.8 Initial conditions

We define the initial distribution of the just-born young over the state-space $(b, s, j) \in R \times \mathscr{S} \times \mathscr{J}$.

[44] Recall that in the model an income of $y = 1$ corresponds to £27,615 in 2015.

We calibrate the density function over such set, $g_0(b,s,j)$ as follows.

From ASHE, we derive the initial skill distribution of the young.

Since individuals are observed at their first job, and we have no information on where they were located until that moment, we proxy the initial location of the young using the distribution of the 18-19 years old population, which captures individuals' locations immediately after graduation from A-levels.

Finally, for the wealth distribution b we take into account how parents often pay for at least part of their kids' homes. Using ... we calculate the joint distribution of location, income, and "help from parents" in purchasing a house, capturing the correlation between skills and inheritance received early enough in life.

C.9 Discount rate ρ

To identify the discount rate, we target the ratio between median wealth and average income in the economy. A higher discount rate is associated with a preference for consumption today and, thus, a lower amount of savings.

From national accounts, we find that the total wealth of British households was 12,572 billions on average during the period 2014 to 2015, of which £1,125 billions were in physical wealth. We exclude the latter, as we do not model durable consumption. During 2014-2015, the average GNI in current prices was 1,876 billions per year. Hence, the average wealth to average income ratio is $11,447/1,876 = 6.10$. From national statistics from ONS, we find that average wealth is 1.94 times the median one. Hence, we use as target the approximate median wealth to average income ratio as $6.10/1.94 = 3.15$.

To calculate total wealth in the model, we sum the value of all owner-occupied housing and all net liquid savings b.

C.10 Amenities and moving shocks and costs

Amenities, moving shocks, and costs are jointly identified by targeting residence shares across locations and relocation flows.

Consider Equation (8). The probability of observing an individual moving from location j to location j' at the appropriate shock age depends on the value $V_t^H(\cdot)$ at destination j', which includes amenities $\mu_{j'}$, moving costs $\kappa_{jj'}$, and the scale of the location taste shock ν. These parameter affect location choice in different ways. To see how, consider the following reasoning.

Since both κ and μ always enter the location choice probabilities as divided by ν, the level of κ and μ is not identified. Instead, we can only identify its value relative to ν. Hence, we consider the normalised parameters ν, $\kappa\nu^{-1}$, $\mu\nu^{-1}$.

Larger ν increase the size of the average idiosyncratic shock, making individuals move more "randomly" relative to the common fundamentals. In particular, notice how for $\nu \to 0$ and $\kappa \to 0$, all individuals would move to the V_t^H-maximising location at every shock age. Moreover, all individuals with identical wealth, skill and age must share the same set of V_t^H-maximising locations j', since there is no idiosyncrasy in taste shocks. In equilibrium, households with identical characteristics would inhabit in identical proportion the locations part of their V_t^H-maximising set. As ν increases, individuals' moves become more "random", and we observe individuals with similar initial characteristics $\Omega = (b,h,j,a,s)$ perform different moves. Hence, ν affects the *variance* of location choices for identical individuals.

Larger $\kappa\nu^{-1}$ lead to less location moves overall. Since moving costs tend to keep individuals in the same location as the one where they started, we will observe less overall moves. Hence, $\kappa\nu^{-1}$ affect the overall *share* of location movers in the economy.

Finally, $\mu_j\nu^{-1}$ allows to pin down the residual variation in moving patterns and residence shares not explained by ν and κ. In particular, consider two locations j and j' with identical wages, house prices, learning opportunities, but which may differ in amenities. Suppose they face the same moving costs from location j'': $\kappa_{j''j} = \kappa_{j''j'}$. Then, if we observe more individuals with state Ω moving from j'' to j rather than j', it must be that j has is more attractive due to its (normalised) amenity component $\mu_j\nu^{-1} > \mu_{j'}\nu^{-1}$.

Hence, by matching the size (moving probabilities vs staying) and distribution of location moves, we can pin down ν, κ, and the amenities vector μ. Since amenities are identified up to a constant, we normalise the amenity level of the first location to $\mu_1\nu^{-1} = 0$.

In practice, we use the degrees of freedom provided by μ_j to perfectly match the residence shares in an inner cycle of our Simulated Method of Moments, given all other parameters. This choice has two advantages. First, when the number of locations is large, it provides large speed gains relative to having to estimate $J-1$ derivatives (one for each location), requiring between $J-1$ and $2(J-1)$ partial equilibrium estimations. As long as the inner loop uses less than such number of iterations to converge, or the overall outer-loop algorithm converges faster, the estimation time decreases. Second, we avoid (directly) estimating μ_j using information from other moments, as it represents a parameter aimed to capture the unexplained part of location choices.

C.11 Preference for house ownership χ

We identify χ by perfectly matching, given all other parameters, the share of individuals who own a house. We do so in an inner loop, together with amenities μ_j.

From the housing survey, we calculate the share of individuals who own a house by weighting each household's tenure (rent or ownership) by the number of adults in the household.

Unfortunately, the UK does not provide data to compute individual-level homeownership. To account for how some young adults tend to live with parents, we perform the following approximations. First, from ONS Population Survey data, we find that approximately 14% of individuals aged 25-34 still live with their parents. We assume these individuals do not own their parents' house. Second, we exclude kids from the household weighting by exploiting information on household composition. In particular, for households that are reported to be "couples with kids", we impose a weight of 2; households that are reported to be "lone parent with kids", we impose a weight of 1. All other households are weighted by the number of adults in the household. This weighting scheme allows to also capture young individuals co-sharing an accommodation. Third, since in the model there is no immigration nor death until the final age, we estimate model-consistent aggregate ownership rates by giving each age an identical weight.[45]

These adjustments lead to an average homeownership rate of 67.4%, ranging from 34.2% for 25-34 years old to 80.0% for 65-79 yers old. The average figure is larger than ONS census data (64% for 2011, 62% for 2021) for two reasons. First, the Census does not weight households by size. However, individuals living together as a couple are more likely to be in the 40-80 y.o. categories, which are also the ones with the highest homeownership. Second, our transformation over-weights older households, as there is no death nor shrinking fertility rates over time (which have slightly reduced the average size of the cohorts)

[45] Practically, the 45-64 and 65+ groups are given double weight relative to the 25-34 and 35-44 due to representing twice-as-large age ranges.

or immigration (with immigrants arriving later in their working life being considerably less likely to own a house relative to similarly-aged peers).

C.12 Warm Glow Parameter θ_0

We identify the warm glow parameter θ_0 by matching the share of homeownership among 65+ individuals. The rationale is that, for $\theta_0 \to 0$, the old would sell all their assets by the time they approach death, as they gain no utility from dying with a positive asset position, including housing. Conversely, if the old care a lot about giving wealth to the next generations, they may avoid buying houses to avoid paying housing transaction costs.

C.13 Labour productivity

We identify labour Z_j productivity by perfectly matching wages. Inverting Equation 13 we obtain the following expression for productivities:

$$Z_j = \frac{w_j}{\alpha} L_j^{1-\alpha}. \tag{36}$$

Both w_j and α are known from the data. Specifically, w_j is the location-specific wage component measured from ASHE and taken as given in the calibration of the other parameters in the Indirect Inference algorithm, while $\alpha = 0.61$ is the labour share in the UK GDP, obtained from ONS national account data. L_j is found integrating the skill supply over the non-retired population, as detailed in Equation 10. Hence, given all other parameters from the first part of the Indirect Inference algorithm, applying the formula in Equation (36) yields the estimates for $\{Z_j\}_{j \in J}$.

C.14 Housing productivity

We identify the steady-state housing productivity Z_{hj} by perfectly matching house prices. Following Guaitoli (2025) and Greaney et al. (2025), we assume that the production function of housing is such that, when the housing market clears, equilibrium house prices are determined as:

$$P_{jt} = Z_{hj,t}(\gamma_j) \left(\frac{H_t^D}{H_{ss}^D} \right)^{\gamma_j},$$

where H_{ss}^D a reference level of housing demand,[46] H_t^D is the housing demanda at time t, and γ_j is a demand elasticity parameter. Since in steady state the term in parenthesis is equal to 1, we simply estimate

$$Z_{hj,ss} = P_j.$$

We postpone the discussion regarding the functional nature of $Z_{hj,t}(\gamma_j)$ to when we perform housing counterfactuals.

[46] We choose to normalise at the steady state, but could be arbitrary number. See Guaitoli (2025).

D Estimation Algorithm

We use the finite difference method to discretise and approximate our continuous-time model.

D.1 Discretisation

We discretise the grid of (potentially) continuous variables (b,h,s,a) as follows:

1. We impose a range $[-20,30]$ for liquid assets b (approximately representing a maximum borrowing limit of £553,000, and a maximum saving limit of £830,000), with an exponential grid of $nB = 46$ points.[47]

2. We discretise house ownership into a set of 6 possible sizes, plus "no ownership" ($h = 0$), for a total of 7 grid points. We call the discretised grid $\mathscr{H}$. The ownership grid is linear over the range $[H_{\min}, H_{\max}]$.

3. We discretise skills using 13 grid points, with a minimum of 0.5 (equivalent to a gross income in the average-productivity location equal to UK's 2015 minimum wage) and a maximum of 4.0, approximately equal to £110,000 gross wage in the average-productivity location. We call the discretised grid $\mathscr{S}$.

4. We impose a range of $a \in [23,80]$, so that the number of shock ages is $nA = 57$, plus a terminal age. We call the discretised grid $\mathscr{A}$, and the age-step between two grid points dt, which we set equal to 1 (hence, each "birthday" corresponds to a shock age). We choose 80 as the terminal age due to the average life expectancy at birth in Britain being around 80 years old.

D.2 Value Function Estimation

We estimate the value function over the discretised grid of state variables $(b,h,j,s,a) \in \mathscr{B} \times \mathscr{H} \times J \times \mathscr{S} \times \mathscr{A}$.

We implement the estimation using the following steps:

1. Using the grid of all possible states (b,h,j,s), for a given age, we estimate the terminal value of the value function, $V(b,h,j,s,\overline{A})$ following the terminal condition:

$$V(b,h,j,s,\overline{A}) = \theta_0 \frac{(\theta_1 + b + (1-\Psi)hp_{jt})^{1-\gamma}}{1-\gamma}.$$

2. We then iterate backwards over the age grid $\mathscr{A}$, starting from $a(k) = \overline{A} - dt$, with $k = nA-1,\ldots,1$. Notice that at each iteration k for age $a(k)$ we know the value function at age $a(k+1) = a(k) + dt$, as it was determined either externally (through the terminal condition at $\overline{A}$) or in the previous iteration of this loop ($a(k+1)$). Hence, we can apply the finite difference method to find the value function at $a(k)$.

   (a) Calculate the forward (f) and backward (b) difference for $\partial_b^x V(b,h,s,j,a(k+1))$, with $x \in \{b,f\}$, using the finite difference method. For $n \in \{1,\ldots,nB\}$ and a generic age, each

[47] Notice that, since we are modelling individuals, this formulation would correspond to a generous £1.1m maximum mortgage, and a £1.6m maximum household liquid asset, plus (see below) £1.3m real estate assets in the most expensive location. This grid is thus sufficient to cover the net wealth of 99% of British households.

forward and backward differences at non-constrained points are defined as:

$$\partial_b^f V(b(n),h,s,j,a) = \frac{V(b(n+1),h,s,j,a) - V(b(n),h,s,j,a)}{b(n+1)-b(n)},\ \forall n \in \mathscr{B}$$
$$\partial_b^b V(b(n),h,s,j,a) = \frac{V(b(n),h,s,j,a) - V(b(n-1),h,s,j,a)}{b(n)-b(n-1)},\ \forall n \in \mathscr{B}$$

At constrained points, we apply the following definition of derivatives.

i. Define a constrained grid point in the forward difference as a point where further accumulation of wealth is not allowed. This corresponds to the wealth point $b(nB)$, since there is no other limit to wealth accumulation at other wealth points, and all points where consumption would be negative with non-negative savings. That is, all states such that $y + q(b)b - (\delta+\tau)hp_j - r_j h^r \leq 0$.

ii. Define a constrained grid point in the backward difference as a point where further borrowing is not allowed. Notice that this corresponds to the sets $\mathscr{C}(h,s,j,a)$, defined as:

$$\mathscr{C}(h,s,j,a) = \left\{n \in \{1,\ldots,nB\} : n = 1 \vee b(n) \leq M_p h p_j \vee b(n) \leq -M_y y(s,j,a)\right\}$$

iii. Using the concept of viscosity solution, notice that, at the constrained points, we must have that the backward-derivative of the value function must be larger or equal to the derivative of the utility function. The reason is that, if it was lower, then we would rather borrow more (to increase consumption) than to save more. However, since we are at a borrowing-constrained point, this must not be the case, as wealth needs to either increase or remain stationary. Hence, the constraints (wealth limit, non-negative consumption, and borrowing constraints) can be implemented by imposing the following conditions:

$$\partial_b^b V(b(n),h,s,j,a) \geq \partial_b u(c,\mathbf{h}).$$

A similar argument, but with opposite signs, goes through for the forward-difference $\partial_b^f V(b(n),h,s,j,a)$.

iv. Computationally, we implement the viscosity solution concept by defining the derivatives as follows:

$$\partial_b^b V(b(n),h,s,j,a) = \partial_b u(c,\mathbf{h}), \forall n \in \mathscr{C}(h,s,j,a),$$
$$\partial_b^f V(b(n),h,s,j,a) = \partial_b u(c,\mathbf{h}), \forall n \in \{nB, n : y + q(b(n))b(n) - \delta h p_j - r_j h^r \leq 0\}.$$

(b) Calculate the optimal consumption $c^x(b,h,j,s,a)$ and rental housing policy function $h^{r,x}(b,0,j,s,a)$ in the forward (x=f) and backward (x=b) cases by solving the following system of equations:

$$\begin{cases} \partial_c^x u(c,\chi h) = \partial_b V^x(b,h,j,s,a) & \text{if } h > 0, \\ \partial_c^x u(c,h^{r*}(c)) = \partial_b V^x(b,0,j,s,a) & \text{if } h = 0, \end{cases}$$

where $h^{r*}(c)$ is the optimal h^r for any given level of consumption c and budget constraint $y + q(b)b \leq c + r_j h^r$, following Equation (**??**). Calculate the optimal consumption and housing

choice also in the wealth-invariant case (x=0) where $\dot{b} = 0$, such that:

$$\begin{cases} c^0 = y(s,j,a) + q(b)b - \delta h, & \text{if } h > 0 \\ c^0 = y(s,j,a) + q(b)b - \eta r_j^{\sigma} (y(s,j,a) + q(b)b)^{1-\sigma}, & \text{if } h = 0 \\ h^{r,0} = \frac{y(s,j,a)+q(b)b-c^0}{r_j}, & \text{if } h = 0 \end{cases}$$

(c) Use the resulting $c^x(b,h,j,s,a)$ and $h^{r,x}(b,0,j,s,a)$ to calculate the backward and forward saving policy functions

$$\begin{cases} \dot{b}^x = y(j,s,a) + q(b)b - c^x(b,0,j,s,a) - r_j h^{r,x}(b,0,j,s,a) & \text{if } h = 0 \\ \dot{b}^x = y(j,s,a) + q(b)b - c^x(b,h,j,s,a) - \delta p_j h & \text{if } h \neq 0 \end{cases}$$

The wealth-invariant case implies $\dot{b}^0 = 0$ by construction.

(d) Solve the finite-difference problem for age a by solving the system

$$\rho \mathbf{v}^k = \mathbf{u}(\mathbf{v}^{k+1}) + \mathbf{A}(k+1)\mathbf{v}^k + \frac{\mathbf{v}^{k+1} - \mathbf{v}^k}{da},$$

where $\mathbf{A}(k+1)$ is the finite-difference transition matrix, featuring only elements on the central, first upper and first lower diagonal and zeros otherwise. Each $(n, n+1)$ element is different from zero only if $\dot{b}^f > 0$; each (n,n) is different from zero only if $-\max \dot{b}^f, 0 + \min \dot{b}^b, 0 \neq 0$, and each $(n-1,n)$ entry is different from zero only if $\dot{b}^b < 0$. The system can be rapidly solved by dividing the matrix $\mathbf{B} = (da^{-1} + \rho) - \mathbf{A}$, of dimensions $nB \times nB$ by the vector $\mathbf{b} = \mathbf{u} + da^{-1}\mathbf{v}$, of dimensions $nB \times 1$.

(e) In case a(k) is a shock age, perform the following additional steps, using the estimated $\mathbf{v}^k$:

i. Estimate the optimal housing decision h' and post-purchase wealth b' for an individual approaching this decision stage with state variables $(b,h,j,s,a(k))$. Notice that h' is the arg max between (i) keeping the current house and wealth, and (ii) selling your house with proceedings $(1-\psi)p_j h$ and buying h' at cost $p_j h'$.[48] Hence, the value function at this stage of life - immediately after the housing tenure choice and before solving the consumption-saving problem - is

$$V(b',h',j,s,a) = \max\{V(b,h,j,s,a), \max_{h'}\{V(b'(h'),h',j,s,a)\}\},$$
$$b'(h') = b + (1-\psi)p_j h - p_j h'.$$

ii. Estimate the expected value function for an individual approaching the location choice moment of the shock age. This corresponds to a simple discrete choice problem of an individua l facing moving costs $\kappa(j,j')$ and Gumbell-distributed idiosyncratic shocks with scale v. Hence, the expected value at this point in time is

$$V(b,h,j,s,a) = v \ln\left(\sum_{j'} \exp\left(v^{-1}(V(b,h,j',s,a) - \kappa(j,j'))\right)\right).$$

Computationally, we implement this log of sum of exponentials using a numerically-stable routine.

[48] Notice that the cost of renting is already included in the value function at $h = 0$, it does not directly go into the initial wealth of the consumption-saving decision, as rental consumption is determined afterwards.

iii. Finally, individuals that reach a shock age experience - before all other shocks and decisions - a skill shock. Immediately before this shock is realised, an individual with state variables (b,h,j,s,a) will have an expected utility function equal to

$$V(b,h,j,s,a) = \int V(b,h,j,s',a) f(s,s') ds',$$
$$\ln(s') = \theta_0 \ln(s) + e_j(\ln(s),a) + \varepsilon_z,\ \varepsilon \sim N(0,\cdot).$$

Here we computationally approximate the integral by calculating the probability mass near each skill gridpoint, and then estimating the integral as (probability-weighted) of the tensor storing the value function, along the skill dimension.

(f) Save the resulting $\mathbf{v}^k$, at the position for age $a(k)$. Repeat the loop from point (2.a).

3. Once all ages have been estimated ($k = 1$), estimate the distribution of skills across space. Starting from age 0 through the last age:
   (a) Calculate the distribution of location choices, starting from the initial condition for age a. The initial condition is either the "birth" one (at age 0), or the age joint distribution of location, skill, housing, and wealth of an individual who is just approaching age a. That is, the joint distribution with CDF $F_a(j,s,h,b) = \lim_{k \to a} F_k(j,s,h,b)$.
   (b) Calculate the distribution of housing choices, using as initial condition the joint distribution of location, skills, housing and wealth observed after the location choice
   (c) Using the Kolmogorov Forward Equations, using as initial condition the distribution of location, skills, housing and wealth observed after the housing choice, calculate the drift of the distribution of liquid wealth between age a and $a + da$.
   (d) Repeat the previous points for every successive age, using as initial conditions the "end distribution" at $a + da$ of the previous age step.
4. Given prices, the skill accumulation process is consistent with the equilibrium distribution of states $g(\Omega)$. Notice that the learning function $e_j(s,a)$ depends on j also due to the average skill in the location $\bar{s}_j$, which is an equilibrium object. Using the updated estimates for the average skill in each location, repeat points (2) and (3) until convergence of the average skill in each location $\{\bar{s}_j\}_{j \in J}$ and the equilibrium density $g(\Omega)$.

D.3 Indirect Inference Algorithm

We perform indirect inference to find the remaining set of parameters $\phi^* \in \mathbb{F}$, for $\mathbb{F}$ being the hyperset of admissible parameters, which minimises a loss function $\mathscr{L}(\phi)$:

$$\phi^* = \arg\min_{\phi \in \mathbb{F}} \mathscr{L}(\phi) = \arg\min_{\phi \in \mathbb{F}} \sum_{n \in N} \omega_n f_n(\hat{X}_n(\phi), X_n)^2,$$

where $\hat{X}_n(\phi)$ represent the simulated moment n from solving the parameter using ϕ, while X_n is the data moments. ω_n represents weights applied to moment n. The function $f_n(\cdot)$ is a function applied to the parameters of family n. We mostly choose $f_n(\cdot)$ to be the percentage deviation between model and data moments, apart from: i) absolute values for moments related to homeownership and wealth, and ii) asinh for moving-related probabilities, due to the fact that using percentage deviations would lead the algorithm to trying to match very small location-to-location moving probabilities, rather than larger ones.

Since amenities μ_j and the preference for homeownership χ mostly affect the population shares and average homeownership, but not other moments, we split the parameter set into two groups. The first

one, which we name Φ, contains all parameters besides μ_j and χ. The second one contains the latter ones, so that $\phi = \{\Phi, \mu, \chi\}$.

To find the loss-minimising set of parameters, we use a gradient descent algorithm:

1. Start from a guess ϕ^k of the parameters. In practice, since all parameters besides κ^a are either strictly or weakly positive, we guess and update their log-values.[49]

2. Use the algorithm described in Section D.2 to estimate the steady state given the guess ϕ^k distribution of population and homeownership. Then, check if the model's population and average homeownership match the data up to a tolerance $\varepsilon > 0$. If not, update the guess of μ_j and χ as follows:

$$\mu_j' = \mu_j - \nabla(N_j - N_j^{\text{data}}),$$
$$\chi' = \chi - \nabla\left(\mathbb{E}(h_i > 0) - \mathbb{E}^{\text{data}}(h_i > 0)\right),$$

   where ∇ is an ad-hoc updating parameter. Then, repeat this step.

3. Use a Newton-based method to find a Φ which minimises the loss function. We use a gradient descent method with individual parameter step-size update. Both during the computation of the derivatives and each parameter update $k' = k+1$, repeat step (2). Once the optimal step falls below a threshold $\delta > 0$ for all parameters in Φ, proceed to the next step.

4. Back out location productivities by perfectly matching the local wages, given population size.

D.4 Estimation Time

Our grid comprises 52 locations, 46 wealth levels, 14 skills, 7 housing points (rent, plus 6 ownership levels), and 63 ages. That is, we have a total of 234,416 grid points per age, and nearly 15 million grid points in total. Location transition grids comprise 2,704 location pairs for a total of 750 million grid points.

We implement the algorithm in Julia. The algorithm estimates the value function and Kolmogorov Forward Equations in 4.2 seconds on a Ryzen 5950x 16-core, 64GB RAM desktop, when single-threaded and 1.5 seconds when multi-threaded.[50] Since learning depends on peers' skills, estimating the steady state requires iterating multiple time the value function and density estimation, for a total computation time of approximately 30-60 seconds (unless $\gamma_2 = \gamma_4 = 0$, in which case the algorithm converges in one single iteration) in single-thread and 8-20 seconds in multi-threading for the Partial Equilibrium steady state.

To compare our algorithm's speed to other benchmarks, a toy version of the model with 5 skills and 52 locations estimates the value function and density matrices in 1.6 seconds (0.8s in multi-threading), and one with 5 skills and 5 locations in 0.15 seconds (0.08s in multi-threading). Notice how the runtime scales less than proportionally in the size of the skill grid $\mathbb{S}$ since, when old individuals do not work and state pension is identical for everybody, the choices of retired individuals do not depend on skills.

[49] This allows to implement a well-behaved and regularised gradient descent. Estimating the log of the parameters is common practice in the Newton-based estimation of economic models, as it allows to remove the scale dependency of gradients on the economic interpretation of different parameters.

[50] Single-threading is necessary when estimating the parameters via parallel gradient descent, as mixing worker parallelisation and multi-threading is not safe on most machines. Multi-threading should be regarded as the benchmark for Equilibrium and transition calculations, taking the parameter as given.

E Additional Material on Counterfactuals

E.1 Welfare calculations

To find the appropriate percentage change x in the value of the consumption bundle $c^{\sigma}\mathbf{h}^{1-\sigma}$, which we call $x = x(b,h,j,s)$, notice that the instantaneous utility function between shock ages, when applying the consumption increase at state $\Omega' = (b',h',j',s',a)$, is

$$U^{eq}(\Omega') = \mu_j \frac{\left((1+x(b,h,j,s))c^{\sigma}\mathbf{h}^{1-\sigma}\right)^{1-1/\theta}}{1-1/\theta}.$$

Importantly, since $x(b,h,j,s)$ is chosen ex-ante relative to the state of the agent at age $a > 0$, the instantaneous utility is always scaled – in every period and state – by a factor $(1+x)^{1-1/\theta}$. A similar relationship holds for the terminal utility level. Considering as the "consumption bundle" the inheritance b plus normalising factor θ_1, we have that

$$v^{eq}(b) = (1+x)^{1-1/\theta} v(b).$$

This intuition carries over at shock ages. When choosing housing, the agents simply choose the highest-value function option among the ones available. Since each of them is simply the continuation value of different housing options, we have that

$$V_t^{H,eq}(\Omega) = (1+x)^{1-1/\theta} V_t^{H}(\Omega).$$

Similarly, right before choosing a location, notice that the log-of-exp utility form implies that

$$V_t^{R,eq}(\Omega) = (1+x)^{1-1/\theta} V_t^{R}(\Omega).$$

Hence, in a similar fashion, the value right before the skill shock realisation is simply

$$V_t^{eq}(\Omega) = (1+x)^{1-1/\theta} V_t(\Omega).$$

Calling $V_t'(b,h,j,s,0)$ the value function at the beginning of the life of an agent in the new steady state, we can use the relationships derived above to derive the closed-form solution for the consumption-equivalent increase as

$$x(b,h,j,s) = \left(\frac{V_t'(b,h,j,s,0)}{V_t(b,h,j,s,0)}\right)^{\frac{\theta}{\theta-1}} - 1.$$

Finally, assuming that the initial distribution of the age-zero population does not change in the new steady state (which is indeed the case as long as $F(b,h,j,s,0)$ is given), the percentage change in welfare is given by integrating the consumption-equivalent $x(\Omega)$ over the distribution of agents of age zero across all states.

$$\begin{aligned}\frac{\Delta W}{W} &= \int (1+x(\Omega))dF(\Omega) - 1 \\ &= \int \left(\frac{V_t'(b,h,j,s,0)}{V_t(b,h,j,s,0)}\right)^{\frac{\theta}{\theta-1}} dF(\Omega) - 1\end{aligned}$$