Housing Prices, Internal Migration, and Intergenerational Mobility*

Qingyuan Chai†

Job Market Paper

(Click here for the most recent version)

September 30, 2024

Abstract

This paper examines how housing affordability affects internal migration in China, with varying impacts across workers from different parental backgrounds. I find that rising housing costs deter migration, with a more pronounced effect on adult children from disadvantaged families. Consequently, these children earn lower incomes than the otherwise similar children from more affluent backgrounds, thereby reducing intergenerational mobility. These findings align with a migration decision model in which housing costs place a disproportionate burden on individuals from less affluent backgrounds. To address the endogeneity of housing prices, I leverage the Housing Purchase Restriction (HPR) policy as a natural experiment and employ an instrumental variable (IV) approach. This policy, which limited the number of properties households could purchase in selected prefectures, reduced housing demand and slowed price growth. A more structural approach allows me to distinguish among destinations and evaluate the effects of various housing policies. The results highlight that the impact of housing costs varies depending on the nature of the prefecture. Providing rent subsidies to migrants to the megacities increases migration among advantaged children more than among disadvantaged children, while a policy targeted at disadvantaged children helps increase their migration and enhance intergenerational mobility.

*I am indebted to my committee members, Kevin Lang, Daniele Paserman, and Raymond Fisman, for their guidance and encouragement. For detailed comments and suggestions, I also thank Jérôme Adda, Huiyao Chen, Oren Danieli, Hanming Fang, Martin Fiszbein, Adam Guren, Douglas Gollin, Joseph Gyourko, Asim Ijaz Khwaja, Patrick Kline, Robert E.B. Lucas, Benjamin Marx, Dilip Mookherjee, Yuhei Miyauchi, Sangyoon Park, and Maisy Wong for helpful comments and discussions. All errors are my own.

†PhD Candidate, Department of Economics, Boston University, email: qchai@bu.edu

1 Introduction

In China, as in many countries including the United States, housing prices have risen faster in the high-salary prefectures than elsewhere. These high housing costs may differentially affect the migration decisions of individuals from more and less advantaged backgrounds, which can exacerbate income disparities and decrease intergenerational mobility. In this paper, using an instrumental variable approach, I find that higher housing costs disproportionately deter migration among adult children from low-education families, even after controlling for the child's own educational attainment. This heterogeneous impact leads to lower earnings for these children compared to those with more educated fathers, thereby reducing economic mobility across generations.

The structural approach reveals a more nuanced pattern: high housing costs in Tier-1 prefectures mainly deter migration among children of highly educated fathers, while for non-Tier-1 prefectures, the negative impact is larger for children of less-educated fathers. These findings suggest that to promote intergenerational mobility, policies should either target workers from disadvantaged backgrounds or reduce housing costs in non-Tier-1 prefectures. My policy experiments confirm this conclusion.

Although there is extensive literature on intergenerational mobility, its determinants, particularly in developing countries, remain unclear (Genicot et al., 2024). Most studies focus on correlations, with few investigating causal factors. Yet, identifying the determinants is crucial for designing effective public policies. This paper moves beyond descriptive associations and provides causal evidence that rising housing prices impede intergenerational mobility. This finding is particularly relevant in light of the recent surge in housing prices in many other countries such as India and the United States (Mahadevia et al., 2012; Ganong and Shoag, 2017).[1]

It is ex-ante ambiguous whether adult children from more or less advantaged backgrounds are more responsive to housing price changes. High housing costs may disproportionately limit migration for children from disadvantaged backgrounds, who often face credit constraints (see, e.g., Cai, 2020), compared to their more affluent peers, thereby reinforcing existing disparities. Conversely, housing price changes may affect affluent children more, as disadvantaged children are often too far from affording migration for small cost changes to matter. In contrast, children from wealthier families, being closer to the decision threshold, can be more sensitive to price fluctuations.[2]

[1]For anecdotes, see "House prices continue to go through the roof" https://www.economist.com/graphic-detail/2021/08/12/house-prices-continue-to-go-through-the-roof and "Is your rent ever going to fall?" https://www.economist.com/international/2024/05/29/is-your-rent-ever-going-to-fall, etc.

[2]Another possibility is that children from wealthier families may also expect higher housing quality, and rising prices for such homes could translate into higher migration costs for these children, discouraging

The ambiguous effects of housing prices on migration decisions across different parental backgrounds subsequently affect how housing prices shape the transmission of socioeconomic status across generations. In other words, the impact of housing prices on intergenerational mobility is also theoretically unclear. To clarify the conditions under which housing prices may either facilitate or hinder intergenerational mobility by affecting migration, I develop a migration decision model. The empirical analysis then reveals which effect dominates in the data and for which destinations.

In this paper, I use China as a testing ground to examine how real estate shocks influence internal migration and intergenerational mobility. China offers a unique setting for several reasons. First, the country has experienced rapid housing price appreciation over the past few decades (Fang et al., 2016), highlighting the significant impact of real estate fluctuations. Housing affordability has become a critical issue for migrants, who complain that "We have no home where there is work, and no work where there is home." Second, housing prices vary substantially across different prefectures, providing rich variation to analyze its effects. The Housing Purchase Restriction (HPR) policy implemented in some prefectures offers a natural experiment to identify the causal impact of housing price shocks. Finally, China's high return to migration is well-documented, especially for workers from rural areas (see, e.g., Lagakos et al., 2020).

In the empirical analysis, I use fathers' education as a proxy for parental background and replace it with imputed fathers' income and household assets for robustness checks.[3] The reduced-from analyses suggest that higher housing costs reduce migration, with the effect being less pronounced for individuals whose fathers have higher education levels. Increases in housing costs also lead to lower incomes, but this negative impact is again mitigated for those with more educated fathers.

As housing prices are correlated with economic conditions, the OLS results are subject to omitted variable bias. For instance, fathers with higher education levels may be more effective at leveraging their social networks to secure job opportunities for their children in high-salary destination prefectures. Therefore, the coefficients might reflect fathers' differential capacity to facilitate their children's access to employment rather than the heterogeneous impacts of housing prices.

To address this concern, I exploit the Housing Purchase Restriction policy introduced in China around 2010 as a natural experiment to construct instrumental variables for housing prices. By restricting the number of properties that each household or firm could

their migration.

[3]There may be concerns that fathers' education may not be a good proxy for income, especially in rural areas. However, data show a strong correlation between education and both current income and household assets for the birth cohort of the sample's fathers, indicating that fathers' education is a valid proxy for income.

purchase, this policy led to an immediate and sharp decrease in housing demand. Since the policy was launched in only 46 prefectures, its impact can be assessed by analyzing outcome variation before and after the policy's implementation across both the affected and unaffected prefectures. This policy has been widely used as an exogenous shock (Chen et al., 2017; Zhao and Zhang, 2022; Liu et al., 2023; Chen et al., 2023).

Regarding heterogeneity, the results imply that basic education is crucial for overcoming housing cost barriers, with the critical factor being whether the father has any schooling. Junior high education is somewhat beneficial, but beyond that, higher levels of education provide minimal additional advantages.

Further analyses reveal that the effects concentrate on children with rural parents, indicating that housing cost barriers primarily influence rural-to-urban migration. Since rural-urban migration is often linked to non-agricultural employment, this finding suggests that housing prices impact agricultural employment differently depending on parental background, which is confirmed by the data. This finding highlights the interaction between housing costs, structural transformation, and intergenerational mobility.

I then use a more structural approach to analyze policy implications, starting with a discrete choice model that accounts for migration as a selection among multiple options rather than a binary decision. I first verify that high housing costs reduce the attractiveness of a prefecture, with varying effects based on fathers' education levels. I then use counterfactual simulations to evaluate three types of rent subsidy policies: a non-targeted subsidy in Tier-1 prefectures, a targeted subsidy in Tier-1 prefectures for workers from disadvantaged families, and a non-targeted subsidy in non-Tier-1 prefectures.

The findings suggest that to encourage migration among workers from disadvantaged families, policies should either directly target these workers in Tier-1 prefectures or focus on non-Tier-1 prefectures. In contrast to these two policies, non-targeted subsidies in Tier-1 prefectures increase migration among advantaged children more than among disadvantaged children. This is because the costs of migration to megacities are so high that many disadvantaged children are far from being able to afford it, and small reductions in housing costs are insufficient to change their decisions. In contrast, children from wealthier families, being closer to the decision threshold, are more responsive to fluctuations in housing prices.

While the discrete choice model provides clear insights into migration responses to policy changes, it does not account for the general equilibrium effect, where changes in population size lead to subsequent adjustments in housing prices, which in turn influence migration until a new equilibrium is reached. To address this, I integrate a housing market into the discrete choice model, creating a full spatial equilibrium model for policy experiments.

Using counterfactual simulations, I provide back-of-the-envelope calculations of the impacts of different policies on intergenerational mobility and agricultural employment. The results suggest that, to improve intergenerational mobility, a targeted rent subsidy in Tier-1 prefectures for disadvantaged migrants is most effective, compared to non-targeted subsidies in either Tier-1 or non-Tier-1 prefectures. Both targeted subsidies in Tier-1 prefectures and non-targeted subsidies in non-Tier-1 prefectures are similarly effective in reducing agricultural employment.

2 Relation to the Literature

First, this paper contributes to the literature on the barriers to internal migration. Previous research has identified various factors, including specific barriers such as the *Hukou* registration system[4] and the land tenure system, alongside broader issues like credit constraints, imperfect information, friction in job matching, risk aversion, transportation costs, and social constraints (see, for example, Zhao, 1999; Bryan et al., 2014; Angelucci, 2015; Munshi and Rosenzweig, 2016; Bazzi, 2017; Chen et al., 2018; Khanna et al., 2021; Morten and Oliveira, 2024). This study complements the literature about the impacts of housing costs on internal migration.

Related to housing costs and migration, Garriga et al. (2023) build a dynamic spatial equilibrium model to explore the interaction between structural transformation and the housing market, finding that rising house prices hinder migration. In the U.S., Ganong and Shoag (2017) document the lower migration rates of low-skilled workers to high-cost housing areas. This paper demonstrates that the effects of housing affordability on migration vary by parental background, impacting intergenerational mobility. To address the endogeneity of housing prices, I construct an instrumental variable based on a real-world policy. The results also imply interactions between housing costs, intergenerational mobility, and structural transformation.

Closely related to this study, Cai (2020) shows that increased credit access boosts migration in China, especially in low-asset villages. Bazzi (2017) finds that transient positive agricultural income shocks increase emigration, particularly in villages with smaller landholders, while persistent income shocks reduce emigration in more developed regions. The findings highlight the role of wealth heterogeneity. While the two papers focus on village-level wealth measures, this paper explicitly analyzes the het-

[4]The *Hukou* system in China functions as a domestic "passport system". It restricts internal migration by limiting access to social services such as education, healthcare, and housing subsidies in areas outside an individual's registered location.

erogeneous impacts based on individual-level parental backgrounds and examines the effects on intergenerational mobility. Moreover, I propose a new type of heterogeneity across household wealth, demonstrating through model and counterfactual analyses that wealthier households may respond more to reduced migration barriers—a finding not covered in previous literature and with important policy implications.

Secondly, it speaks to the literature that investigates the determinants of intergenerational mobility. While the phenomenon of intergenerational persistence is well-documented, the factors causally driving it are not fully understood. Previous studies highlight the importance of economic shocks, access to education, and segregation (for example, Parman, 2011; Feigenbaum, 2015; Chyn et al., 2022; Biasi, 2023). This paper proposes housing prices as a new determinant. Additionally, existing literature predominantly investigates developed countries (Genicot et al., 2024); this paper focuses on a developing context where geographic inequality and migration play crucial roles in shaping intergenerational mobility.

Close to this study, Ward (2022) provides evidence that migration enhances intergenerational mobility, especially noting its effectiveness for individuals from lower-income families. This paper complements his work by examining a different setting and focusing on the impact of elevated housing prices, a trend recently witnessed in many countries.

Furthermore, this paper connects to the growing literature on the causal relationship between childhood exposure to better neighborhoods and their intergenerational mobility. Specifically, a series of papers examine the impacts of the Moving to Opportunity (MTO) experiment (e.g. Chetty and Hendren, 2018).[5] In contrast, this study focuses on policies that could enhance intergenerational mobility by encouraging migration in adulthood. This distinction underscores different life stages—childhood versus adulthood— where targeted interventions can break the cycle of poverty and dependence on parental backgrounds.

3 Background

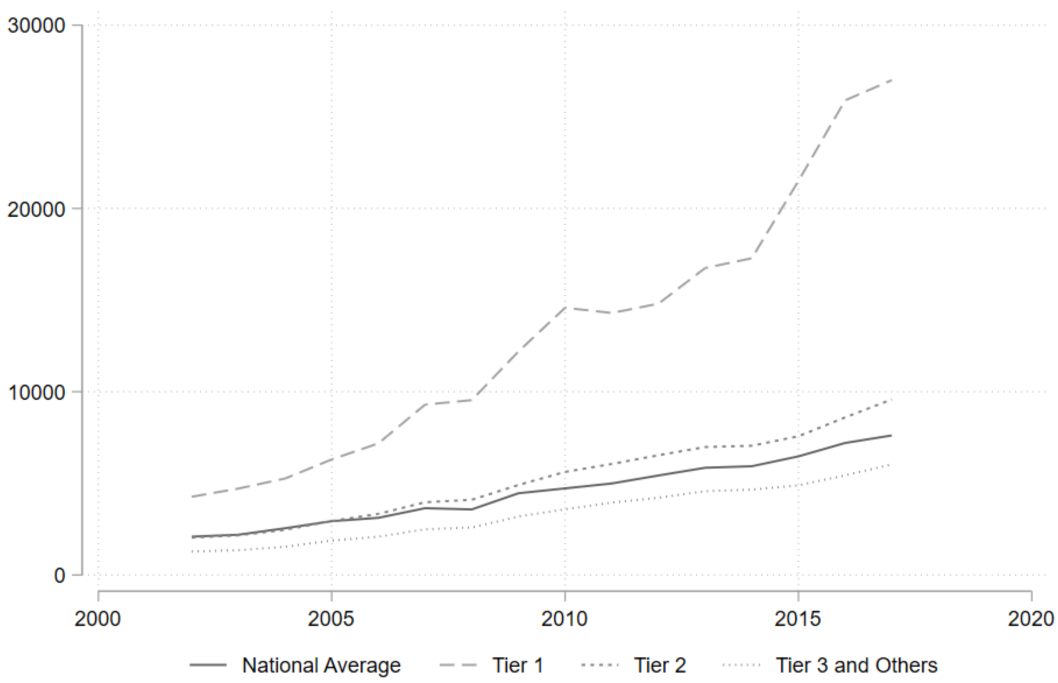
3.1 Housing Boom in China

The 1998 reform, which privatized the housing sector, initiated a rapid growth of the real estate markets in China. As shown in Figure 1, housing prices in China have surged dramatically over the past two decades. Even after accounting for housing quality, from

[5] Similarly, Alesina et al. (2021) find that in Africa, an additional year in a high-mobility region between ages 5 and 12 significantly increases the likelihood of children from uneducated families completing primary school.

2002 to 2013, the average yearly growth rate of residential housing prices in 35 major Chinese prefectures was 11.4 percent (Sun and Zhang, 2020). Such increases have been attributed to factors such as land supply restrictions, increases in mortgage supply due to fiscal stimulus, expected economic growth, etc. (see, for example, Wang and Wen, 2012; Fang et al., 2016; Sun and Zhang, 2020).

Figure 1: Residential Housing Prices by Prefecture Tier (yuan/sq.m)



NOTE: This figure presents the nominal yearly housing prices across prefectures of different tiers from 2002 to 2017. Tier-1 prefectures are comprised of Beijing, Shanghai, Guangzhou, and Shenzhen. Thirty-one prefectures are classified as Tier-2. Other prefectures fall into Tier-3.
SOURCE: Rogoff and Yang (2020)

In tandem with the rise in housing prices, rental costs also surged, a factor of particular concern to migrants. According to data from the China Migrants Dynamic Monitoring Survey (CMDS), the median monthly rent in Tier 1 prefectures doubled from 500 RMB in 2011 to 1000 RMB in 2017, not taking into account the potential decrease in residential quality that migrants might accept to manage higher costs. After adjusting for the Consumer Price Index (CPI), this rise translates to a jump from 403 to 717 in 2000 RMB.

Despite the overall upward trend in housing prices over the past few decades, there are significant temporal and regional disparities across prefectures. This variation can be attributed to factors such as the differential housing supply elasticity, diverse local house purchasing policies, and spatial inequality in economic development (Saiz, 2010; Gyourko et al., 2022). Appendix Figure B.1 illustrates this variation by showing the residualized housing prices, which are derived from regressing housing prices on prefecture fixed effects and year fixed effects.

3.2 Housing Purchase Restriction Policy

On April 17, 2010, the State Council of China released the "Notice of the State Council on Resolutely Curbing the Excessive Rise of House Prices in Some Prefectures" aimed at moderating the surging housing prices in some urban regions. After this notice, Beijing pioneered the implementation of a Housing Purchase Restriction (HPR) policy on May 1, which limited households to buying no more than one new property. Subsequently, this policy was extended to an additional 45 major prefectures across China from late 2010 to early 2011, as detailed in Table B.1 (dates are from Chen et al. (2017)). While the specific regulations varied among these prefectures, the general framework of the HPR policy typically restricted Chinese households to owning no more than two properties, with a mandatory two-year interval between purchases. The policy was in place until 2014, after which it began to be slowly removed.

The policy was adopted abruptly and unexpectedly. Immediately following the implementation of the HPR policy, housing demand declined noticeably, leading to slower price growth and a reduction in the number of houses sold in the affected prefectures (Du and Zhang, 2015; Li et al., 2017). This paper utilizes this slowdown in the housing market, which affects migrants' housing affordability and, thus, their migration decisions.

4 Model

How do housing prices affect migration decisions and intergenerational mobility? To explore the relevant factors, I construct a model of migration decisions. The model indicates that housing price changes can either increase or decrease migration and intergenerational mobility, depending on parameters and specific distributions of disutility regarding migration.

4.1 Setup

In this section, I describe a discrete version of the model for simplicity. A continuous version can be found in Appendix Section A, with similar main results. Assume fathers'

Table 1: Adult Children's Income by Parental Backgrounds and Migration Status

		Y Stay	Migrate
F	H	Y_H^0	Y_H^1
	L	Y_L^0	Y_L^1

socioeconomic status, F, is high (H) or low (L). If the adult child stays, their income is Y_H^0 or Y_L^0, depending on the father's type. If the child migrates, their income is Y_H^1 or Y_L^1. Table 1 shows children's income by parental background and migration status.

The model makes the following assumptions. First, it posits that all children are homogeneous, and it is the father's type that affects the children's income. Second, $Y_L^1 > Y_L^0$ and $Y_H^1 > Y_H^0$, meaning that migration increases income. Third, the model assumes that incomes at the origin and destination are realized after the migration decision. That is, the migration decision is made based on the expected income from either staying or migrating rather than on actual income.

Migration incurs a monetary cost, denoted as $M(h, F)$, which depends on the housing cost h and the father's type F. Assume that $\frac{\partial M(h,F)}{\partial h} > 0$, meaning that an increase in housing costs increases migration costs. In addition to monetary costs, individuals experience psychological disutility associated with migration, denoted by τ. This disutility, τ, is independent of the father's type and the children's income. The cumulative distribution function (CDF) and probability density function (PDF) for τ are denoted as $F_\tau(\tau)$ and $f_\tau(\tau)$, respectively.

Combining the above benefits and costs of migration, a child will migrate if:

$$Y_F^1 - Y_F^0 - M(h, F) > \tau \tag{1}$$

where F represents either H or L.

Denote the left-hand side as $B(h, F)$, which is the net benefit of migrating. For a child faced with father's type F and housing cost h, the probability that they migrate is $Pr(migrate|h, F) = Pr(B(h, F) > \tau) = F_\tau(B(h, F))$.

In the empirical analyses, I regress the indicator for migration on the interaction term between paternal backgrounds and housing costs. As father's type is binary in the model, the sign of the interaction term corresponds to the sign of $\left(\frac{\partial Pr(migrate|h,F)}{\partial h}\bigg|_{F=H} - \frac{\partial Pr(migrate|h,F)}{\partial h}\bigg|_{F=L} \right)$. It can be derived that:

$$\frac{\partial Pr(migrate|h, F)}{\partial h} = -f_\tau(B(h, F))M_F'(h) < 0 \tag{2}$$

and

$$\left.\frac{\partial Pr(migrate \mid h,F)}{\partial h}\right|_{F=H} > \left.\frac{\partial Pr(migrate \mid h,F)}{\partial h}\right|_{F=L} \iff \frac{M'_H(h)}{M'_L(h)} < \frac{f_\tau(B_L)}{f_\tau(B_H)} \quad (3)$$

where B_F denotes $B(h,F)$, $M'_F(h)$ denotes $\frac{\partial M(h,F)}{\partial h}$ which is positive, and $F = H$ or L.

In the empirical analysis, I also examine the effects of housing costs on children's income. According to the model, the expected income of children, regardless of migration status, is given by:

$$\begin{aligned} E[Y|h,F] &= F_\tau(B(h,F))Y_F^1 + [1 - F_\tau(B(h,F))]Y_F^0 \\ &= F_\tau(B(h,F))(Y_F^1 - Y_F^0) + Y_F^0 \end{aligned} \quad (4)$$

The impact of housing costs on income is:

$$\frac{\partial E[Y|h,F]}{\partial h} = -f_\tau(B(h,F))M'_F(h)(Y_F^1 - Y_F^0) < 0 \quad (5)$$

I regress children's income on the interaction term between paternal backgrounds and housing costs. The sign of the interaction term corresponds to the sign of $\left(\left.\frac{\partial E[Y|h,F]}{\partial h}\right|_{F=H} - \left.\frac{\partial E[Y|h,F]}{\partial h}\right|_{F=L}\right)$. It can be derived that:

$$\left.\frac{\partial E[Y|h,F]}{\partial h}\right|_{F=H} > \left.\frac{\partial E[Y|h,F]}{\partial h}\right|_{F=L} \iff \frac{M'_H(h)}{M'_L(h)} < \frac{f_\tau(B_L)}{f_\tau(B_H)}\frac{(Y_L^1 - Y_L^0)}{(Y_H^1 - Y_H^0)} \quad (6)$$

4.2 Discussion

Equation (3) suggests that the impact of housing prices on migration decisions depends on two factors: the heterogeneous costs, $\frac{M'_H(h)}{M'_L(h)}$, and the different position at the disutility distribution, $\frac{f_\tau(B_L)}{f_\tau(B_H)}$. First, for the magnitude of $\frac{M'_H(h)}{M'_L(h)}$, if credit constraints exist, housing costs increase migration costs more for children with L type fathers, namely $M'_F(h)$ decreases in F. For $\frac{f_\tau(B_L)}{f_\tau(B_H)}$, $f_\tau(B_F)$ represents the sensitivity of disutility to the net benefit of migration. In the population, $\frac{f_\tau(B_L)}{f_\tau(B_H)}$ measures the relative number of children at the margin of deciding whether to migrate. If more children with high-type fathers are at the margin and fewer with low-type fathers, $f_\tau(B_L)$ is smaller than $f_\tau(B_H)$, and vice versa.

Equation (6) suggests that the two mechanisms described above, heterogeneous migration costs and the relative number of people at the margin, also apply to the analysis of the heterogeneous impacts of housing costs on children's income. There is an additional term, $\frac{(Y_L^1 - Y_L^0)}{(Y_H^1 - Y_H^0)}$, which measures the relative benefit of migrating between children with L and H type fathers. I assume that parental background holds more value in the hometown, so $\frac{(Y_L^1 - Y_L^0)}{(Y_H^1 - Y_H^0)}$ is larger than one, meaning that migration is more beneficial for children with low-type fathers.[6] Given this assumption, Equation (6) is a necessary condition of Equation (3). The underlying idea is that children from low-income backgrounds benefit more from migration, but they also suffer more when housing costs rise, leading to greater negative impacts on their income.

Overall, the impact of housing costs on migration decisions and children's income varies based on parental background. Whether the impacts are larger for children with high- or low-type fathers is theoretically ambiguous. Empirical analysis is crucial to identify the dominant factors in the data.

5 Empirical Approach

I begin the analyses by exploring how housing prices influence migration decisions and income, with a focus on how these effects vary by fathers' years of education. To address potential endogeneity concerns regarding housing prices, I then employ a 2SLS estimation, using the Housing Purchase Restriction (HPR) policy to construct the instrumental variable.

5.1 Baseline

I run the following regression to examine the heterogeneous impacts of housing prices on migration and income across different parental backgrounds. For adult children i born in year y whose origin is prefecture o, I estimate the following equation:

$$\begin{aligned} Outcome_{ioy} = \beta_1 FatherEduy_{ioy} \times HPgap_{oy} + \beta_2 FatherEduy_{ioy} \\ + \beta_3 HPgap_{oy} + \Pi Z_{oy} + \Omega X_{ioy} + \mu_y + \eta_o + \epsilon_{ioy} \end{aligned} \tag{7}$$

where β_1 is the coefficient of interest. $Outcome_{ioy}$ is either the indicator for migration or log income.[7] Migration is defined as residing in a prefecture different from one's origin, where the origin is defined as the prefecture in which the individual resided at age 14.[8] $HPgap_{oy}$ is the housing price gap between the origin and potential destinations, which I describe in detail below. $FatherEduy_{ioy}$ denotes the individual's father's years of

[6] There is suggestive evidence in the data that supports this assumption, though it is not causal.

[7] One potential concern is that, as there exists a sizeable rural-urban gap in the cost of living, the variation in income may capture mainly the price difference in different areas but not the real difference. To address this concern, I estimate the main results using income adjusted for rural-urban and provincial price differences based on the approach in Brandt and Holz (2006), and the results remain robust.

[8] Results are robust using other ages such as 16 and 18.

education. The individual-level control variables, X_{ioy}, include gender, parents' *Hukou* status, and education fixed effects. The origin-by-birth year level controls, Z_{oy}, include log wage levels, employment, and exports at the origin prefecture, as well as weighted averages of log wage levels, employment, and exports at the destination prefectures.[9] μ_y and η_o represent fixed effects for birth year and origin prefecture, respectively.

Given the persistent rise in housing prices in China over recent decades, there is concern that $FatherEduy \times HPgap$ may largely reflect cohort-specific effects of fathers' education. To address this, I control for the interaction between fathers' education and birth year fixed effects, allowing the impact of fathers' education to vary across birth cohorts. This approach isolates the time-varying effects of education, thereby refining the estimate of β_1.

Although I use cross-sectional data from a single year, the survey provides information on migration history, including origin, destination, and year of migration, allowing me to infer housing prices during working ages. Ideally, each migration decision would be matched with housing market characteristics at the time when the individual was contemplating whether to migrate. In practice, following Sun and Zhang (2020), the expected housing price gap the individuals were exposed to is calculated as follows. The expected housing price gap each individual was exposed to is defined as the price difference between the origin and potential migration destinations, weighted by the importance of each destination for that origin. The time series of the gap is then averaged, with weights based on the significance of each year during key ages when migration decisions are made in the individual's life cycle.

To construct the expected housing price gap, first, using China's 2010 census sample, I calculate the age-specific migration probability, defined as the proportion of individuals who migrated at age k relative to the total number of migrants.[10] Appendix Figure B.3 shows the plot of the age-specific migration probability.

$$a_k = \frac{\#\, of\ migrants\ migrated\ at\ age\ k}{\#\, of\ migrants}$$

To minimize the influence of parental migration decisions, the calculation begins at age 16. It ends at age 45, as data indicates minimal migration occurs after that age. Adjusting these thresholds yields similar results. If an individual aged l in the survey

[9]For the reduced-form results, macroeconomic controls are included for migration outcomes to address potential omitted variable bias but are excluded from the income outcomes to avoid issues related to bad controls. The results remain robust regardless of whether macroeconomic controls are included or not for both outcome variables. A detailed discussion of this issue can be found in Appendix Section C.

[10]I assume the age-specific migration rate is homogeneous across prefectures. The results remain robust when using province-specific rates. The analysis using rates calculated from the 2000 and 2005 census data yields similar results.

year, where $l < 45$, they were exposed no later than age l, in which case l is the upper bound of the summation.

Then, for an individual born in year y, at age l, originating from prefecture o, the expected housing price gap to which they were exposed is defined by the following expression:[11]

$$HPgap_{oy} = \frac{\sum_{k=16}^{min\{l,45\}} a_k \times (lnprice\_dest_{o,y+k} - lnprice_{o,y+k})}{\sum_{k=16}^{min\{l,45\}} a_k}$$

Where $lnprice_{o,y+k}$ is the log housing price in the origin prefecture o in year $(y + k)$, and $lnprice\_dest_{o,y+k}$ is the weighted average housing price at potential destinations for the same year.

To calculate the weighted average housing prices of potential destinations, I use Census 2000 data to generate weights based on migration patterns prior to the sample period.[12] The proportion of migrants from a given origin moving to each specific destination serves as the weight. For instance, if 50% of migrants from prefecture o moved to Shanghai, Shanghai's housing price is given a 50% weight in the calculation for prefecture o. The formula is as follows:

$$lndestprice_{ot} = \sum_p w_{op} lnprice_{pt}$$

where w_{op} is the fraction of migrants from origin prefecture o who migrated to prefecture p, and $\sum_d w_{op} = 1$.[13]

Variation in the housing price gaps arises from two sources: first, children from the same origin but born in different years face different housing price shocks at ages crucial for making migration decisions; second, children from different origins experience different price shocks due to varying origins and the corresponding differently weighted destination prefectures.

Control variables at the origin-by-birth-year level are weighted similarly. For example, the log wage at the origin is calculated as a weighted average based on age-specific migration probabilities. Destination wages are also weighted, first by migration patterns

[11]For transparency and to address concerns about functional form, the results remain robust when using the housing price gap at a specific age, such as 20.

[12]Alternatively, I replace migration pattern weights with the ratio of the destination prefecture's 1999 GDP to the distance from the origin, following the concept of market access as in Donaldson and Hornbeck (2016). The results remain robust.

[13]Instead of using the weighted average of destination housing prices, I create individual-destination level observations to analyze the impact of housing prices at each specific destination on the likelihood of migrating there. The results remain qualitatively similar.

and then by age-specific migration probabilities.

5.2 Instrumental Variable

Housing prices are highly correlated with local economic conditions. To address the endogeneity issue, an instrumental variable is needed to capture the impacts of housing prices. The Housing Purchase Restriction (HPR) policy is leveraged as a natural experiment to construct the instrumental variable.

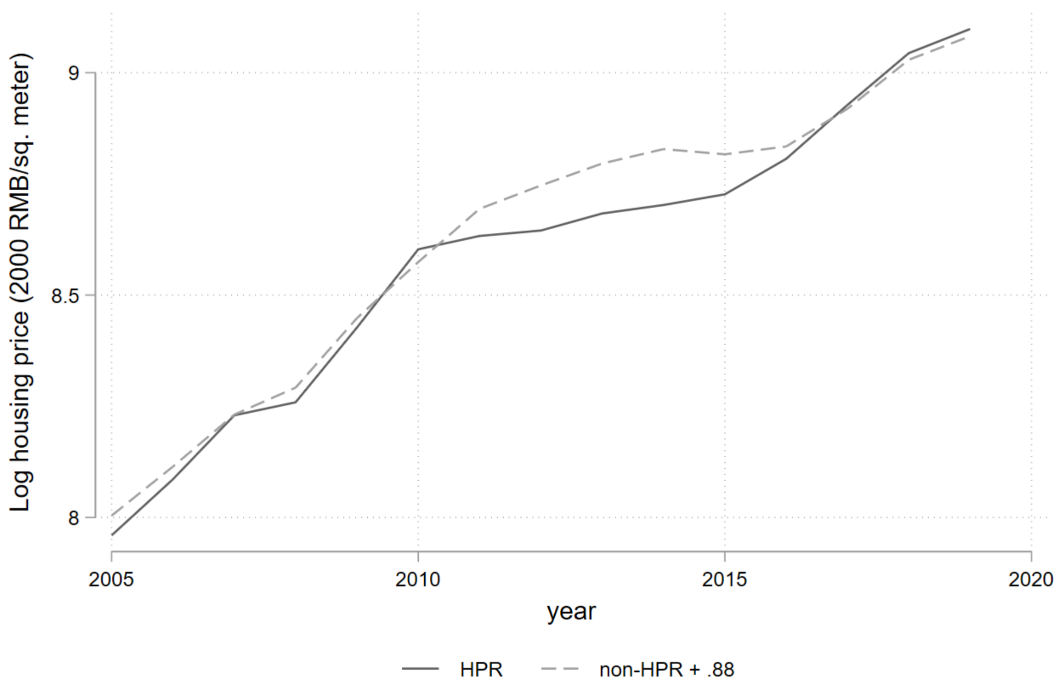


Figure 2: Housing Prices for HPR and non-HPR prefectures

NOTE: The figure shows the housing prices time series between the prefectures that implemented the HPR policy and those that never adopted it, measured by log yuan per square meter adjusted to 2000 RMB. To make the comparison more transparent, the housing prices of the never-treated prefectures are shifted upward by 0.88 log points.

Figure 2 shows the housing prices time series between the prefectures that ever implemented the HPR policy and those that never did, measured in log yuan per square meter, adjusted to 2000 RMB. To facilitate comparison, housing prices in the never-treated prefectures are shifted upward by 0.88 log points. Visually, the pre-trend appears parallel before the implementation of the HPR policy around 2010. Since the implementation of the policy, the growth of the treated prefectures immediately slowed, while the housing prices in untreated prefectures kept growing. As the treated prefectures gradually relaxed the restriction in 2014-2016, the differences returned to the original level.

Appendix Figure B.2 shows the event study plot for the policy. The estimates indicate that the gap in log housing prices between ever-treated and never-treated prefectures was stable before the policy, narrowed significantly during the policy, and returned to the original level after the policy was lifted in most treated prefectures.

To construct the instrumental variable, I first create a prefecture-year dummy indicating whether the HPR policy is in effect. These dummies for the origin and destination prefectures are then weighted and summed in the same manner as the housing price gaps. The resulting weighted average of the policy indicators, HPR_{oy}, serves as the instrumental variable for the housing price gap.

The instrumental variable is essentially a linear shift-share IV, where the shock is exogenous. The validity of this approach has been demonstrated by Borusyak et al. (2022).[14]

6 Data

6.1 Housing Prices and Local Economic Conditions

The panel data on housing prices are obtained from the CEIC database, which provides annual selling prices for residential properties in each prefecture. As the individual level data is in 2017, the housing price data used covers the period from 2000 to 2017. Housing price is measured in log yuan per square meter and adjusted to 2000 RMB using the national Consumer Price Index (CPI).

I use housing prices as proxies for housing costs for migrants. Migrants usually do not purchase houses at the destinations due to both *Hukou* restrictions and credit constraints. According to the CMDS dataset, about 10-20% migrants who migrated across prefectures purchased a house at the destination. What matters more for the migration decision is the rental prices at the destination. Unfortunately, data on rental prices are limited in the Chinese context. Data suggests that rental prices and house purchasing prices are highly correlated.[15] Therefore, I use housing prices to proxy the residential costs faced by migrants.[16]

[14]As the shift-share instrumental variable in this paper is linear in the exogenous shocks, it is valid given that the origin fixed effects are controlled for. I further examine the robustness using the recentering approach following Borusyak and Hull (2023). The results remain robust.

[15]The correlation coefficient between log rent and log housing prices is 0.89, indicating a strong correlation between the two stats. The rent data is sourced from WIND.

[16]There may be concerns that migrants' living conditions are poor, which could make their rent less sensitive to housing prices at their destination. However, using CMDS data, I find that rents paid by

The CEIC database also provides data on local economic conditions, including employment, wage, GDP per capita, exports, FDI, and fiscal expenditures.

6.2 Individual Characteristics

Individual characteristics data are sourced from the China Household Finance Survey (CHFS), a nearly nationally representative survey project executed by the Research Center for China Household Finance.[17] The CHFS dataset provides a comprehensive and detailed description of household economic and financial behaviors. Further information about the dataset can be found in Gan et al. (2014).

This study uses data from the 2017 CHFS. I construct a sample where interviewees are treated as adult children and use their self-reported income and parental information for analysis. In addition to respondents' migration history, the CHFS provides detailed income data, including from the informal sector and farming, which helps mitigate selection biases that commonly arise in studies lacking income data for workers in these sectors. The CHFS also collects basic parental information, such as education and *Hukou* status, regardless of whether parents co-reside with the respondent or are still alive, reducing selection bias typically seen in studies focused only on cohabiting household members.

I restrict the sample as follows: First, since migration behavior may differ for workers from megacities, I exclude observations whose origin is Tier-1 prefectures (Beijing, Shanghai, Shenzhen, and Guangzhou). Second, I limit the sample to individuals aged 17-55. The results are not sensitive to changes in the age range. Appendix Table B.2 shows the descriptive statistics.

7 Reduced-Form Results

7.1 OLS Results

I estimate Equation (7) using OLS, and the results are shown in Column (1) of Table 2. The coefficient for $HPgap$ is negative, indicating that higher housing costs reduce migration. Because fathers' education is demeaned, the coefficient represents the impact of housing prices on children whose fathers have average years of schooling. The positive coefficient for $FatherEduy \times HPgap$ suggests that fathers' education can mitigate the

migrants across different percentiles (5th, 10th, ..., 95th) are all affected by housing prices at their destination.

[17] Tibet and Xinjiang are not included in the dataset. Together, they account for less than 2% of China's population, according to the National Bureau of Statistics.

adverse effects of housing costs. However, this mitigation is insufficient to fully offset the housing cost barrier, which would require approximately 28 years of education—an unachievable amount in this context.

Table 2: OLS Results

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
FatherEduy	0.005***		0.023***	
	(0.001)		(0.004)	
FatherEduy× HPgap	0.009***	0.009***	0.028***	0.027***
	(0.001)	(0.001)	(0.006)	(0.006)
HPgap	-0.252***	-0.275***	-0.795**	-0.764**
	(0.073)	(0.073)	(0.334)	(0.344)
FatherEduy× birthyearFE	-	Y	-	Y
Obs.	14,976	14,976	12,068	12,068
Adj. R-sq	0.424	0.425	0.352	0.353
Mean(Dep. Var.)	0.202	0.202	9.798	9.798

NOTE: This table shows the OLS results. Compared to the odd columns, the even columns control for fathers' years of education interacted with birth year fixed effects, so the level term of fathers' education year is absorbed, and the estimated coefficients are not shown in the table. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and the individual's own education level. For Columns (1) and (2), log wage, log employment, and log export at the origin prefecture, the weighted averages of the log wage, log employment, and log export at the potential destination prefectures are controlled in addition. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

To address concerns that $FatherEduy \times HPgap$ may primarily capture heterogeneous impacts of fathers' education across birth cohorts, I include an interaction between fathers' education and birth year fixed effects. This allows the effect of fathers' education to vary by birth year, thereby isolating the time-varying impact of a father's education. The results are shown in Column (2) of Table 2. The coefficients of interest barely change.

Columns (3) and (4) replace the outcome variable with the log of the adult children's income. The sample sizes are smaller, as only observations with available income information are included. The coefficient of $FatherEduy$ is positive and significant, suggesting that fathers' education is positively correlated with the children's income.[18] The negative coefficients of $HPgap$ indicate that higher housing costs are associated with lower income. This aligns with expectations, as migration typically leads to income

[18]The magnitude is close to the findings in the literature. For example, Lee et al. (2024) find that an exogenous one-year increase in parents' schooling increases children's lifetime earnings by 1.2-1.7 percent on average. My estimates are somewhat larger.

gains, and higher housing costs act as a barrier to migration. Additionally, the positive and significant coefficients for the interaction term $FatherEduy \times HPgap$ indicate that higher levels of fathers' education can mitigate the adverse effects of housing costs on children's income. Column (4) controls for fathers' years of education interacted with birth year fixed effects, and the estimates change little.

7.2 IV Results

Unobserved factors, such as job opportunities, might confound the impacts of housing prices on migration. As a booming economy attracts migrants while higher housing costs deter them, the coefficient of $HPgap$ is likely attenuated in the OLS results. The positive coefficient for $FatherEduy \times HPgap$ could suggest that fathers with higher education levels are more effective at leveraging their social networks to secure job opportunities for their children in destination prefectures. Consequently, the interaction term may capture the varying abilities of fathers to help their children access employment rather than the differential impact of housing prices.

Table 3 shows the results using the instrumental variable approach. Row "K-P F-stat" shows the first-stage *F-stats*, which are larger than 10, suggesting strong first stages. The "LM test" row presents statistics testing whether the matrix of first-stage coefficients has full column rank. High values indicate that the instruments provide sufficient independent variation to explain the endogenous variables. The results of the first stages of Columns (3) and (6) are shown in Appendix Table B.3. The impacts of the HPR policy on housing costs are negative, as the HPR policy reduces housing demand and thus slows housing price growth.

In Columns (1)-(2), the negative coefficients for $HPgap$ suggest that larger housing price gaps reduce migration, while the positive interaction term indicates that fathers' education mitigates the deterring effect of higher housing costs, though not enough to overcome the housing barrier fully. In Columns (4)-(5), where the dependent variable is children's log income, the negative coefficients for $HPgap$ indicate that higher housing price gaps reduce income. Similarly, the positive interaction term shows that fathers' education helps offset the adverse effects of housing costs, but not completely.

Because fathers' education affects children's education, one potential explanation for the findings is that fathers' education proxies children's education, and it is children's education that helps overcome housing cost barriers. While such a channel also suggests intergenerational persistence, the policy implications differ based on whether the focus is on enhancing educational opportunities or improving labor market access. To exclude the effect of children's education, I include the interaction term between the children's

Table 3: IV Results

	(1) Migration	(2) Migration	(3) Migration	(4) lnIncome	(5) lnIncome	(6) lnIncome
FatherEduy× HPgap	0.011***	0.011***	0.009***	0.022**	0.021**	0.027***
	(0.002)	(0.002)	(0.002)	(0.010)	(0.009)	(0.010)
HPgap	-0.753***	-0.779**	-0.699**	-2.852**	-2.720**	-3.009**
	(0.292)	(0.303)	(0.297)	(1.288)	(1.327)	(1.351)
Eduy× HPgap	-	-	Y	-	-	Y
FatherEduy× birthyearFE	-	Y	Y	-	Y	Y
K-P F-stat	40.013	37.685	25.508	66.829	62.223	41.908
LM test	70.140	67.772	68.667	114.214	107.637	110.246
Obs.	14,976	14,976	14,976	12,068	12,068	12,068
Mean(Dep. Var.)	0.202	0.202	0.202	9.798	9.798	9.798

NOTE: This table shows the main results using the HPR policy to construct the instrumental variable. Fathers' years of education interacted with birth year fixed effects are controlled for in Columns (2), (3), (5) and (6). In Columns (1), (2), (4), and (5), the endogenous variables are $HPgap$ and $FatherEduy \times HPgap$; the instrumental variables are HPR and $FatherEduy \times HPR$. In Columns (3) and (6), the endogenous variables are $HPgap$, $FatherEduy \times HPgap$, and $Eduy \times HPgap$; the instrumental variables are HPR, $FatherEduy \times HPR$, and $Eduy \times HPR$. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and education. For Columns (1)-(3), log wage, log employment, and log export at the origin prefecture, the weighted averages of the log wage, log employment, and log export at the potential destination prefectures are controlled in addition. The K-P F-stat row shows the F-stats of the first stages. The LM test row shows the statistics for the underidentification test. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

own education and the housing price gap, $Eduy \times HPgap$, in Columns (3) and (6). The coefficients for $FatherEduy \times HPgap$ remain similar, indicating that fathers' education continues to help through channels other than children's education.

To interpret the magnitudes of the coefficients, it's important first to understand the scale of the housing price gap. The housing price gap is a weighted average of the difference between the log housing prices at potential destinations and the origin. The mean of the housing price gap is 0.55, indicating that housing prices at the destinations are approximately 173% of those at the origin.

I then calculate the impacts of a 0.1 change in the housing price gap. Since both fathers' education and the housing price gap are demeaned, the magnitudes of the coefficients reflect impacts at the mean housing price gap and mean fathers' education. For the effect of housing costs on migration, the coefficient of -0.699 in Column (3) implies that a 10 percent increase in the housing price gap decreases migration by about 7 percentage points, which is roughly 35% of the mean migration rate.[19] The coefficient

[19]For comparison, Bazzi (2017) find that a 10 percent increase in cumulative rainfall shocks raises migration rates by 4 percentage points, about 35% of the mean rate. Cai (2020) find that a village banking program increases migration by 19% of the baseline mean in control villages.

for $FatherEduy \times HPgap$ in Column (6) suggests that a 0.1 increase in the housing price gap could raise intergenerational persistence by 12%.[20] For comparison, Feigenbaum (2015) finds that a downturn during the Great Depression that is one standard deviation more severe increases intergenerational persistence by 39%. Chyn et al. (2022) find that a 1 standard deviation increase in racial segregation reduces average mobility by 17% for children of parents at the 1st income percentile, where the impact is the largest.

7.3 Robustness

7.3.1 Measurement of Intergenerational Mobility

A large body of literature addresses the measurement of intergenerational mobility (IGM). The most common measures (of low IGM) include the intergenerational income elasticity (IGE), obtained by regressing children's log income on parents' log income, the intergenerational income correlation (IGC), and the rank-rank correlation, which regresses children's percentile rank on their parents' percentile rank (see, e.g., Genicot et al., 2024, for a survey of mobility measures). In this paper, I use the impact of fathers' education on children's income as a measure of intergenerational persistence, as fathers' education is reported best. Education is also less affected by transitory fluctuations and remains fixed once completed, unlike income, making it a common proxy for parental lifetime income (Solon, 1992; Gong et al., 2012). To address concerns like attenuation bias discussed in the literature, I conduct the following robustness checks.

First, since the data is cross-sectional, children's income is from the year 2017. Income data from a single year may be subject to transitory shocks or measurement errors. However, this is less concerning here, as income is the dependent variable, not the independent variable. While measurement error in the dependent variable may add noise, it does not introduce bias. The independent variable, fathers' education, is stable and less prone to transitory shocks or measurement errors. To further address this concern, I follow Nybom and Stuhler (2017) and calculate average income across the 2015, 2017, and 2019 waves of the CHFS, provided income data are available. Appendix Table B.7 presents the results using multi-year average income. The estimated coefficients are very similar between single-year and multi-year income measures.

Secondly, since my dataset includes children from various prefectures and birth years, the same level of fathers' education may represent different socioeconomic backgrounds. For example, a father with a high school education in a remote area may have a very different status compared to one in Beijing. To address this concern, I use residualized

[20] Using the coefficient for $FatherEduy$ in Column (4), 0.023, the calculation for the change in intergenerational persistence due to a 0.1 change in $HPgap$ is 0.027 × 0.1 / 0.023 = 12%.

fathers' years of education, obtained by regressing fathers' education on *Hukou* status and origin-by-birth-year fixed effects. The results, shown in Appendix Table B.8, are similar to the main results in Table 3.

While education is an important indicator of parental background, it does not capture all aspects. Fathers with the same educational level may still have very different incomes and household wealth. To address this, I re-estimate the main results using imputed fathers' income or household wealth. I impute fathers' lifetime income based on parental education, *Hukou* status, Communist Party membership, job position, and coastal region indicators. These variables are relatively stable over the life cycle and correlate strongly with lifetime earnings, reducing lifecycle bias. Additionally, using estimated rather than actual income helps mitigate attenuation bias from temporary income fluctuations (Fan et al., 2021). The details of the imputation procedure are described in Appendix Section D.

Table 4: Robustness: Impute Fathers' Income

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
lnFatherIncome× HPgap	0.185***	0.138***	0.355**	0.437***
	(0.034)	(0.037)	(0.150)	(0.165)
HPgap	-0.868***	-0.770***	-2.775**	-3.069**
	(0.304)	(0.297)	(1.299)	(1.330)
Eduy× HPgap		0.008***		-0.014
		(0.003)		(0.013)
lnFatherIncome× birthyearFE	Y	Y	Y	Y
K-P F-stat	37.911	25.658	63.563	42.697
LM test	67.733	68.631	108.382	111.236
Obs.	14,955	14,955	12,051	12,051
Mean(Dep. Var.)	0.202	0.202	9.798	9.798

NOTE: This table shows the results using imputed fathers' income in place of fathers' education. Imputed fathers' incomes interacted with birth year fixed effects are controlled in all columns. In Columns (1) and (3), the endogenous variables are $HPgap$ and $FatherIncome \times HPgap$; the instrumental variables are HPR and $FatherIncome \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherIncome \times HPgap$, and $Eduy \times HPgap$; the instrumental variables are HPR, $FatherIncome \times HPR$, and $Eduy \times HPR$. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and education. For Columns (1) and (2), log wage, employment, and exports at the origin prefecture, weighted averages of the log wage, employment, and exports at the potential destination prefectures are controlled in addition. The K-P F-stat row shows the F-stats of the first stages. The LM test row shows the statistics for the underidentification test. A rejection of the null indicates that the matrix of first-stage coefficients has full column rank, i.e., the model is identified. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

I re-estimate the main results using imputed fathers' income instead of fathers' education. Table 4 presents the findings. The F-statistics for the first stage remain above 10, indicating strong first-stage regressions. While the results follow similar patterns to the main table, the magnitudes differ due to the use of different measures of parental background.

A similar procedure is applied to impute household assets. Appendix Table B.10 shows the results where fathers' education is replaced with imputed household assets. The results remain robust.

7.3.2 Additional Robustness Tests

The housing price gap is at the origin times birth year level, where the corresponding weights of the potential destinations are specific to each origin. There may be concerns that some unobservable economic conditions at the origin can confound the results. To address this concern, I include origin × birth year fixed effects, which would capture the time-variant characteristics in the origin and the origin-specific weighted averages of destination shocks. The results are shown in Appendix Table B.4. The estimated coefficients of interest remain close to the main results.

Recent econometric literature (Feigenberg et al., 2023) highlights that omitted variable bias may confound interaction terms, even when level terms of confounders are controlled. To address this, I control for the interaction between fathers' education and potential confounders. In Appendix Table B.4, I include the interactions between fathers' education and variables such as employment, wages, fiscal expenditure, exports, and FDI at both the origin and destination prefectures, along with the level terms for these economic conditions. The results are similar to the main results in Table 3.

Characteristics of the origin prefecture may influence migration responses to housing price changes. For instance, individuals from economically developed or popular destination prefectures may exhibit different migration behaviors. To address this, I exclude from the sample the top ten prefectures by GDP per capita or the top ten most popular destinations, then rerun the main regressions.[21] Additionally, people born in prefectures adjacent to megacities such as Beijing or Shanghai may choose to live in origin and commute rather than rent or purchase houses in the megacities. Another concern is that housing prices may have geographical spillovers, which could affect migrations through an income effect. To check whether these concerns matter for my result, I run the main regressions excluding the prefectures that are adjacent to the Tier-1 prefectures: Beijing,

[21] The top ten prefectures by GDP per capita in 2000 were Shenzhen, Shanghai, Zhuhai, Wuxi, Suzhou, Guangzhou, Beijing, Xiamen, Dongying, and Hangzhou. Aside from megacities like Beijing and Shanghai, the top ten vary across years. Using other years yields similar results. Based on 2010 census data, the top ten destinations are Beijing, Shanghai, Shenzhen, Guangzhou, Dongguan, Hangzhou, Suzhou, Chengdu, Ningbo, and Foshan.

Shanghai, Shenzhen, or Guangzhou.[22] Appendix Table B.5 shows the results restricting the origin prefectures as described above. The results are close to the main findings.

The housing price data used in the main results comes from the CEIC database. To verify that the observed patterns are not specific to CEIC, I also use data from Anjuke, a leading online real estate platform in China which is known for its extensive listings and comprehensive services for property buying, selling, and renting. The results are shown in Appendix Table B.6. The estimates are close to the main results using the CEIC data, suggesting that I am not capturing dataset-specific patterns.

7.4 Heterogeneous Analysis

7.4.1 Fathers' Education Levels

This section explores the heterogeneous impacts across different subgroups. First, I analyze which level of fathers' education matters most in overcoming housing barriers.

Table 5 shows the results of interacting different levels of fathers' education with housing price gaps. Fathers' education is categorized into four levels: 1) no schooling (32.36% of the sample), 2) primary school (32.91%), 3) junior high school (19.21%), and 4) senior high or above (15.52%), with 'no schooling' as the reference group. To assess the incremental impact of each level, I use indicators for reaching or exceeding specific thresholds. For example, the coefficient of $1FatherEduLevel \geq 3 \times HPgap$ reflects the added impact of a junior high education compared to primary school in overcoming housing barriers. A significant positive coefficient suggests an advantage from higher education, while an insignificant one indicates that further education beyond the previous level does not significantly help.

Columns (1) and (3) of Table 5 show the impacts on migration. The results reveal that, compared to fathers with no schooling, having a primary school education significantly help overcome housing barriers, as indicated by the positive and statistically significant coefficients. The magnitudes of the coefficients of having a junior high school education are nonnegligible, although noisily estimated. Further education beyond junior high school, such as senior high school or higher, does not yield statistically significant additional benefits in mitigating the challenges posed by housing expenses. Columns (2) and (4) display the impacts on income, which are consistent with the migration findings, suggesting that the critical threshold for overcoming the adverse effects of housing prices on income lies at the basic education level.

[22] The adjacent prefectures are Baoding, Shaoguan, Suzhou, Zhongshan, Tianjin, Dongguan, Zhangjiakou, Chengde, Jiaxing, Qingyuan, Langfang, Foshan, Huizhou.

Table 5: Heterogeneity: Fathers' Education Levels

	OLS		IV	
	(1)	(2)	(3)	(4)
	Migration	lnIncome	Migration	lnIncome
$1\{\text{FatherEduLevel} \geq 2\} \times$ HPgap	0.045***	0.130*	0.073***	0.205*
	(0.011)	(0.069)	(0.019)	(0.115)
$1\{\text{FatherEduLevel} \geq 3\} \times$ HPgap	0.027*	0.092	0.018	0.132
	(0.015)	(0.069)	(0.022)	(0.090)
$1\{\text{FatherEduLevel} \geq 4\} \times$ HPgap	-0.000	-0.020	0.009	-0.036
	(0.019)	(0.076)	(0.025)	(0.102)
HPgap	-0.297***	-0.811**	-0.668**	-2.863**
	(0.073)	(0.347)	(0.303)	(1.363)
K-P F-stat			14.462	23.837
LM test			65.974	108.881
Obs.	14,976	12,068	14,976	12,068
Mean(Dep. Var.)	0.202	9.798	0.202	9.798

NOTE: This table shows the IV results for heterogeneity across children with parents of different education levels. The interaction term between the individual's education and the housing price gap is included in all columns. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and education. For Columns (1) and (3), log wage, employment, and exports at the origin prefecture, weighted averages of the log wage, employment, and exports at the potential destination prefectures are controlled in addition. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

The findings imply that fathers with low education attainment, who likely have limited access to credit, are much less able to help their children overcome housing barriers, thereby limiting the children's opportunities for migration and economic advancement. These results align with the broader literature on poverty traps and credit constraints, which emphasize how such constraints severely restrict people's ability to migrate, even within their own country. These findings underscore the importance of affordable housing policies, particularly for those at the bottom of the income distribution, as such measures can greatly enhance economic mobility and reduce persistent poverty.

7.4.2 Rural vs Urban

I also examine heterogeneity between children from rural and urban areas.[23] Columns (1) and (2) of Table 6 present results for rural children, while Columns (3) and (4) show results for urban children. The findings indicate that parental background plays a

[23]Children with at least one parent holding a rural *Hukou* are classified as from rural areas. I avoid using children's *Hukou* type directly, as it correlates with migration behavior.

significant role in helping rural children overcome housing cost barriers. In contrast, for urban children, the coefficients are smaller and not statistically significant, suggesting a negligible effect of parental education on housing barriers. The differences between the rural and urban groups are statistically significant, highlighting distinct impacts of housing cost barriers across the two groups.

Table 6: Heterogeneity: Rural vs Urban

	Rural		Urban	
	(1) Migration	(2) lnIncome	(3) Migration	(4) lnIncome
FatherEduy× HPgap	0.012***	0.044**	0.006	-0.013
	(0.003)	(0.014)	(0.004)	(0.014)
K-P F-stat	20.126	38.658	16.983	17.990
LM test	56.090	103.368	50.227	52.676
Obs.	10,081	8,303	4,895	3,765
Mean(Dep. Var.)	0.210	9.516	0.179	10.419

NOTE: This table shows the IV results for heterogeneity across children from rural versus urban areas. The interaction term between the individual's education and the housing price gap is included in all columns. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and education. For Columns (1) and (3), log wage, employment, and exports at the origin prefecture, weighted averages of the log wage, employment, and exports at the potential destination prefectures are controlled in addition. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

These findings highlight the challenges rural families face and underscore the critical role of parental background in mitigating these barriers for workers from rural areas. Given that most rural workers migrate from rural to urban areas, the results suggest that rural-urban migration is the most affected by housing cost barriers compared to other types of migration.

7.5 Impacts on Structural Transformation

In the heterogeneity analysis, I find that housing costs have a greater impact on individuals from rural areas. Since rural workers' migration is often tied to non-agricultural employment, housing prices may affect agricultural employment differently based on parental background. This variation underscores a complex interaction between housing costs, structural transformation, and intergenerational mobility. To explore this further, I examine the effects of housing costs on agricultural employment.

Table 7: Impacts on Agricultural Employment

	Rural Sample	
	(1)	(2)
FatherEduy× HPgap	-0.006*	-0.008**
	(0.003)	(0.003)
HPgap	1.031*	1.105*
	(0.601)	(0.613)
Eduy× HPgap	-	Y
FatherEduy× birthyearFE	Y	Y
K-P F-stat	30.197	20.063
LM test	56.791	56.562
Obs.	8,123	8,123
Mean(Dep. Var.)	0.293	0.293

NOTE: This table shows the results using agricultural employment as the outcome. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and education, log wage, employment, and exports at the origin prefecture, weighted averages of the log wage, employment, and exports at the potential destination prefectures. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table 7 shows the results of this analysis, where the outcome variable is whether the individual works in the agricultural sector. The sample is restricted to children from rural areas. The findings suggest that higher housing costs increase the likelihood of working in the agricultural sector, as shown by the positive and significant coefficients for $HPgap$. Moreover, the interaction between fathers' education and housing costs reveals that higher levels of fathers' education mitigate the impact of housing costs on the probability of working in agriculture. The negative and significant coefficients for the interaction term $FatherEduy \times HPgap$ indicate that as fathers' education increases, the influence of housing costs on agricultural employment decreases. This suggests that families with more educated fathers are better equipped to navigate the challenges posed by housing costs, enabling their children to pursue non-agricultural employment opportunities.

The heterogeneous effects of housing costs on agricultural employment across different parental backgrounds highlight the role of housing costs in shaping structural transformation and social mobility. The findings suggest that housing costs may reinforce existing inequalities and perpetuate poverty by trapping workers in agriculture. This interaction between housing costs and agricultural employment underscores the need for policies that improve housing affordability, particularly for rural populations, to promote more inclusive economic growth.

8 Discrete Choice Model

In the reduced-form analyses, I aggregate the impacts of housing prices at potential destinations into a weighted average, treating migration as a binary decision—whether to migrate or not. However, in reality, migration is not a binary choice but a selection from multiple potential destinations, where the housing price at one destination can influence migration flows to other destinations.A discrete choice model captures this complexity by modeling migration decisions as discrete choices.

The discrete choice model can then be used to simulate the outcomes of various policy interventions, such as housing subsidies. Additionally, it allows me to assess how policy impacts vary based on targeted destination prefectures and population subgroups. By comparing simulated outcomes for different policies, the model offers insights into effective strategies for increasing migration through improved housing affordability. Additionally, it enables the evaluation of how different housing policies can reduce intergenerational persistence and barriers to structural transformation.

8.1 Setup

Since the discrete choice model generates the labor supply and housing demand for the full spatial equilibrium model in Section 9, I provide a more detailed setup in this section.

Each worker chooses to live in the prefecture that offers the most desirable combination of wages, housing prices, and moving costs. A worker residing in prefecture p supplies one unit of labor and earns a wage W_p. The worker consumes housing M at a local price of rH_p, where r is the discount factor converting purchase prices to rental prices, and a national good O with a normalized price of one. The worker maximizes Cobb-Douglas preferences for housing and the national good, subject to a budget constraint:

$$\begin{aligned} \max_{M,O} \quad & \ln\left(M^{\zeta}\right) + \ln\left(O^{1-\zeta}\right) \\ \text{s.t.} \quad & O + rH_pM \leq W_p. \end{aligned}$$

Workers' relative preference for the national good versus local housing is determined by ζ, where $0 \leq \zeta \leq 1$, and is assumed to be constant across all workers. The worker's optimized utility from consumption can be expressed as the indirect utility function, V_p, for living in prefecture p:

$$V_p = ln(W_p) - \zeta ln(rH_p) = w_p - \zeta h_p - \zeta lnr$$

where $w_p = ln(W_p)$ and $h_p = ln(H_p)$. The worker's optimized utility function also determines their housing demand, D_p:

$$D_p = \frac{\zeta W_p}{r H_p} \tag{8}$$

Additionally, worker i from origin prefecture o incurs financial and psychological moving costs, C_{ifop}, when relocating to destination prefecture p:

$$\begin{aligned} C_{ifop} = &\beta_f^h h_p + \beta^d lnDistance_{op} + \beta^r ruralHukou_i \times \mathbf{1}\{o \neq p\} \\ &+ \beta^m male_i \times \mathbf{1}\{o \neq p\} + \pi_f + \xi_{fp} \end{aligned}$$

The term $\beta_f^h h_p$ represents the transient residential cost incurred when a worker searches for a job upon arriving at the destination. This impact varies by parental background, f, as workers from affluent families are better equipped to cover these upfront costs. $\beta^d lnDistance_{op}$ captures transportation costs or the disutility of being far from home, which depends on the log distance between the origin and destination prefecture. A worker can also choose to stay in the origin, in which case the log distance is set to zero. β^r represents the differential cost for rural workers when migrating; they often benefit more than urban workers due to significantly improved job opportunities and amenities at the destination. β^m represents the differential migration cost between men and women. π_f is the fixed effects capturing taste differences among workers with different parental backgrounds. The residual, ξ_{fp}, captures the impacts of unobserved destination-specific characteristics, such as local amenities. Since housing prices can be correlated with local amenities and thus endogenous, I instrument housing prices using the HPR policy as in the main analysis.

Each worker has an idiosyncratic preference for each prefecture, represented by ϵ_{ifop}, drawn from a Type I Extreme Value distribution. To simplify notation and discussion of estimation, I normalize the utility function by dividing each worker's utility by the standard deviation of ϵ_{ifop}, so that ϵ_{ifop} follows a standard Type I Extreme Value distribution. For consistency and clarity, with a slight abuse of notation, I use the same notation for the parameters in the normalized utility function as in the previous unnormalized utility function expressed in wage units. β^w is a new parameter that represents the utility derived from wages. The resulting utility for a worker choosing prefecture p is:[24]

$$U_{ifop} = \beta^w (w_p - \zeta h_p) + \beta^h_f h_p + \beta^d lnDistance_{op} + \beta^r ruralHukou_i \times \mathbf{1}\{o \neq p\}$$
$$+ \beta^m male_i \times \mathbf{1}\{o \neq p\} + \pi_f + \xi_{fp} + \epsilon_{ifop}$$

I rewrite this equation as:

$$U_{ifop} = \delta_{fp} + \chi_{iop} + \epsilon_{ifop}$$

where $\delta_{fp} = \beta^w (w_p - \zeta h_p) + \beta^h_f h_p + \pi_f + \xi_{fp}$ and $\chi_{iop} = \beta^d lnDistance_{op} + \beta^r ruralHukou_i \times \mathbf{1}\{o \neq p\} + \beta^m male_i \times \mathbf{1}\{o \neq p\}$. $lnDistance_{op}$, $ruralHukou_i$, and w_p are taken as exogenous, while h_p is treated as endogenous. A worker can choose to live in the origin, in which case $o = p$.

This setup is the conditional logit model, first formulated in this utility maximization context by McFadden (1973). The probability that worker i choose to migrate to prefecture p is:

$$Pr_{ifop} = \frac{\exp\left(\delta_{fp} + \chi_{iop}\right)}{\sum_{n \in N} \exp\left(\delta_{fn} + \chi_{ion}\right)}$$

where N is the set of all prefectures.

8.2 Estimations

In practice, I utilize panel data across different prefectures and years, allowing me to exploit both cross-sectional and temporal variation for estimation. Specifically, the utility obtained by a worker born in year y from origin o with parental background f when choosing to reside in destination prefecture p is:

$$U_{ifopy} = \beta^w w_{py} + \beta^h_f h_{py} + \beta^d lnDistance_{op} + \beta^r ruralHukou_i \times \mathbf{1}\{o \neq p\}$$
$$+ \beta^m male_i \times \mathbf{1}\{o \neq p\} + \pi_{fy} + \mu_p + \xi_{fpy} + \epsilon_{ifopy}$$

where some variables are allowed to vary across birth cohorts, and time-invariant prefecture fixed effects μ_p are included. I combine $-\beta^w \zeta h_{py}$ and $\beta^h_f h_{py}$ into $\beta^h_f h_{py}$, as they cannot be separately identified in estimation.

The magnitudes of the coefficients on wages and housing prices reflect the elasticity of workers' demand for a small prefecture with respect to local wages and housing prices, given the functional form assumption.[25]

[24] $-\zeta lnr$ is dropped as it is a constant.

[25] Given the assumed distribution of workers' idiosyncratic preferences for prefectures, the elasticity of demand for workers with father's education f and birth cohort y for prefecture p with respect to local

The probability that worker i choose to migrate to prefecture p is:

$$Pr_{ifopy} = \frac{\exp\left(\delta_{fpy} + \chi_{iop}\right)}{\sum_{n\in N} \exp\left(\delta_{fny} + \chi_{ion}\right)} \tag{9}$$

where N is the set of all destination prefectures.

I estimate the model in two steps. First, I estimate Equation (9) using MLE. Secondly, I take the estimated $\hat{\delta}_{fpy}$ and apply a 2SLS estimation utilizing the implementation of HPR policy as the instrument for housing prices h_{py} to estimate β^h_y.

In practice, there is not enough data to estimate δ_{fpy} for each birth year and each level of fathers' education. Instead, I divide the sample into early and late birth cohorts based on birth years and categorize fathers' education into high and low, i.e., both f and y are binary. This approach allows me to estimate four δ's for each prefecture. Appendix Figure B.4 shows the goodness of fit of the first step. I then use these estimate δ's to estimate the following equation using a 2SLS approach:

$$\begin{aligned}\hat{\delta}_{fpy} = {} & \beta_1 FatherEduHigh_f \times HousingPrices_{py} + \beta_2 HousingPrices_{py} \\ & + M'_{py}\Theta + \mu_p + \pi_{fy} + \xi_{fpy}\end{aligned} \tag{10}$$

where $FatherEduHigh_f$ is an indicator for fathers' education being high type. M_{py} are macroeconomic conditions, which consist of employment rate, GDP per capita, and log wage. μ_p are prefecture fixed effects, and π_{fy} are fathers' education times birth cohort fixed effects. ξ_{fpy} is the residual. Housing prices are instrumented using the HPR policy, and the interaction term $FatherEduHigh \times HousingPrices$ is instrumented by $FatherEduHigh \times HPR$.[26]

The results of the second step are shown in Table 8. Columns (1) and (2) show the OLS estimates without and with macroeconomic controls, respectively. Columns (3) and (4) display the IV estimates, also with and without controls. The negative and significant coefficients for housing prices indicate that higher housing costs reduce a worker's utility of residing in a prefecture. The positive and significant coefficients for the interaction between fathers' education and housing prices suggest that higher paternal education partially offsets the adverse impact of housing prices, though not enough to counterbalance the negative effects of high housing costs fully.

housing prices is $(1 - s_{fopy})\beta^h_f$, where s_{fopy} denotes the share of workers choosing to live in prefecture p. For a small prefecture, where s_{fopy} is close to zero, the demand elasticity for rent is approximately β^h_f.

[26]The regressions assume that destination incomes equally impact utility for workers regardless of parental background, as supported by the data. Interaction terms between macroeconomic conditions and parental background are insignificant and have small coefficients when included in the regression.

Table 8: Conditional Logit

	OLS		IV	
	(1)	(2)	(3)	(4)
FatherEduHigh× HousingPrices	0.380***	0.384***	0.747**	0.758**
	(0.107)	(0.110)	(0.322)	(0.326)
HousingPrices	-1.181***	-1.011***	-4.050***	-3.606**
	(0.347)	(0.376)	(1.230)	(1.436)
FatherEduHigh× birthcohort FE	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Controls	-	Y	-	Y
K-P F-stat			12.880	7.772
LM test			19.100	9.569
Obs.	607	603	607	603
Mean(Dep. Var.)	0.178	0.180	0.178	0.180

NOTE: Standard errors are clustered at the prefecture level. Columns (1) and (2) show the OLS estimates, both without and with macroeconomic controls, which consist of employment rate, GDP per capita, log wage, and log export. Columns (3) and (4) show the IV estimates, again with and without these controls.

The coefficients on housing prices are consistent with the existing literature. Column (4) estimates show that the elasticity of workers' demand for a small prefecture with respect to local housing prices is approximately -3.606 for workers with low-education fathers and -2.848 for those with high-education fathers. In comparison, Diamond (2016) reports an elasticity of around -2.496 for non-college workers and -1.312 for college-educated workers, though her estimates reflect elasticity with respect to rent rather than housing purchase prices.

The results of the first stage of the 2SLS estimation are shown in Appendix Table B.11. The coefficients have the expected signs. Specifically, the coefficients of the HPR policy, when housing price is the independent variable, are negative, indicating that the implementation of the HPR policy negatively affects housing prices.

The results remain robust under more complex specifications, such as incorporating workers' education, including existing migration patterns between origin and destination, adjusting thresholds for fathers' education and birth cohorts, and allowing the effects of distance and rural *Hukou* to vary by parental background.

8.3 Counterfactuals

I begin by examining policies targeting Tier-1 prefectures, as they are the primary migration destinations. Approximately one-third of all migrants relocate to these areas.

However, this overall figure masks significant heterogeneity based on parental education: about one-quarter of children with low-education fathers move to Tier-1 prefectures, compared to 41% of those with high-education fathers.

I first analyze the impact of a housing price regulation policy that offers migrants a 10% rent subsidy in Tier-1 prefectures. In other words, migrants face housing prices that are 90% of the actual prices in Tier-1 prefectures. I then compare the migration behavior of children from different parental backgrounds before and after the policy implementation. Panel A of Table 9 shows the percentage point changes in migration behavior resulting from the policy. The interpretation is that, for example, the proportion of workers with low-education fathers staying in their origin prefecture decreases from 85.15% to 83.98%, a change of -1.17 percentage points. The results imply that, while the fraction of workers with low-education fathers who migrate to Tier-1 prefectures increases by 1.45 percentage points, the increase is much larger for those with high-education fathers.

Table 9: Counterfactuals: Discrete Choice Model

	Change in Percentage Points (%)		
FatherEdu	Stay	Migrate to Tier-1	Migrate to Non-Tier-1
Panel A: 10% Rent Subsidy in Tier-1 Prefectures			
L	-1.17	1.45	-0.28
H	-1.76	2.41	-0.66
Panel B: 10% Rent Subsidy in Tier-1 Prefectures for Migrants with Low-Education Fathers			
L	-1.17	1.45	-0.28
H	0.00	0.00	0.00
Panel C: 10% Rent Subsidy in Non-Tier-1 Prefectures			
L	-3.72	-0.26	3.98
H	-2.65	-0.64	3.29

NOTE: This table shows the percentage point changes in migration behavior among workers from different parental backgrounds for each policy. For example, the -1.17 percentage point change in Panel A indicates the reduction in the proportion of workers with low-education fathers staying in their origin prefecture, a change from 85.15% to 83.98%.

The previous results have shown that increases in housing costs decrease the utility more for children with disadvantaged backgrounds. Why is migration to Tier-1 prefectures more responsive among the advantaged compared to the less advantaged? A key factor is that the high initial housing prices in Tier-1 prefectures make workers with disadvantaged backgrounds unlikely to consider migrating there. Marginal changes in housing costs are not large enough to affect their migration decisions. In contrast, workers with high-education fathers are more likely to be at the margin of making migration decisions. This is consistent with the model in Section 4, which discusses how the heterogeneous impacts of housing prices across parental backgrounds depend on both varying migration costs and the relative number of children at the margin across these groups.

Additional tests suggest that the larger distances to Tier-1 prefectures also contribute to fewer children with low-type fathers being at the margin compared to those with high-type fathers. However, merely reducing the distance difference between the two groups is insufficient to make children with low-type fathers respond more to the rent subsidy than their high-type counterparts.

The results above indicate that a more targeted approach may be more effective in encouraging migration among children from disadvantaged families. I thus evaluate the impact of a targeted housing subsidy policy that offers a 10% discount in Tier-1 prefectures specifically for migrants whose fathers have low education levels. Under this policy, migrants with low-education fathers face housing prices at 90% of the actual value, while those with highly educated fathers pay the total price. Table 9 Panel B shows the results. This targeted approach significantly increases migration among workers with low-education fathers.

No response is observed for workers with high-education fathers because they face the same housing prices as before. In reality, housing prices in Tier-1 prefectures will change endogenously with the influx of workers with low-education fathers. We take this general equilibrium effect into account in the next section where we estimate the full spatial equilibrium model.

An alternative policy is to provide rent subsidies in non-Tier-1 prefectures, which might require lower fiscal expenditure due to their relatively lower rents compared to Tier-1 prefectures.[27] Under this policy, migrants would face housing prices at 90% of the actual prices in non-Tier-1 prefectures. Panel C of Table 9 presents the results. Unlike subsidies in Tier-1 prefectures, subsidies in non-Tier-1 prefectures have a greater impact on the migration behavior of workers with low-education fathers compared to those with high-education fathers. For both groups, the impacts are significantly larger than the previous policies due to the broader coverage of destination prefectures.

Overall, the three counterfactual analyses indicate that policy design should consider the nature of the different destination prefectures. In Tier-1 prefectures, a targeted

[27]A non-targeted policy also reduces the risk of workers' misreporting parental backgrounds, which is difficult to verify.

approach is more effective in increasing migration among workers from disadvantaged families, whereas in non-Tier-1 prefectures, targeting is not necessary.

9 Spatial Equilibrium Model

The above analyses focus on the household side without addressing the effects of migration on destination prefectures. However, migration undoubtedly has general equilibrium effects. For instance, a decline in housing prices in a prefecture attracts more migrants, which in turn drives up housing prices due to increased demand, limiting further migration. This section incorporates these general equilibrium effects and simulates the counterfactual outcomes for the three policies discussed in the previous section. Moreover, back-of-the-envelope calculations are provided to assess the impacts on intergenerational persistence, agricultural employment, and the associated fiscal expenditure.

The spatial equilibrium model follows a standard framework with one tweak: housing price impacts vary by parental background. It extends the discrete choice model from the previous section by incorporating a housing market that adjusts housing prices in response to changes in population size.

9.1 Setup

9.1.1 Labor Demand

I assume that firms in each prefecture produce a homogeneous tradable good using identical constant returns to scale technology. Thus, firm-level labor demand directly reflects prefecture-level aggregate labor demand. Each prefecture has a representative profit-maximizing firm, with output prices normalized to one. In each prefecture p for birth cohort y, the representative firm takes productivity A_{py} as given and produces output using the production function:

$$Y_{py} = A_{py}L_{py}$$

where L stands for labor in the prefecture. Since there are a large number of firms and no barriers to entry, the labor market is perfectly competitive, and firms hire such that wages equal the marginal product of labor. Namely $w_{py} = lnA_{py}$. Wages are treated as exogenous and remain unaffected by changes in the prefecture's population size.

9.1.2 Housing Supply

Following Diamond (2016), housing production depends on construction materials and land, with developers acting as price-takers and selling homogenous houses at marginal cost:

$$h_{py} = MC(\kappa_{py}, \eta_{py})$$

where $MC(\kappa_{py}, \eta_{py})$ is the marginal cost function based on local construction costs κ_{py} and land costs η_{py}. Land cost η_{py} is a function of aggregate housing demand. The log housing supply equation is parameterized as follows:

$$h_{py} = \ln(\kappa_{py}) + \gamma \ln(D_{py}) \tag{11}$$

Here, γ is the elasticity of housing prices with respect to demand, which is calibrated from the literature. $\ln(\kappa_{py})$ is unobserved and included in the residuals.

9.1.3 Equilibrium

Equilibrium in this model is defined by a menu of wages and housing prices $\{w_{py}, h_{py}\}$ with populations $\{L_{py}\}$, such that:

- The labor demand equals labor supply:

The labor demand function is given by:

$$w_{py} = lnA_{py}$$

Labor supply is the total expected population of prefecture p for birth cohort y, calculated as the sum of the probabilities that each worker chooses to reside in this prefecture.

$$L_{py} = \sum_{i \in L_y} \frac{\exp\left(\delta_{fpy} + \chi_{iop}\right)}{\sum_{n \in N} \exp\left(\delta_{fny} + \chi_{ion}\right)} \tag{12}$$

where L_y represents the set of workers in birth cohort y across the nation. $\delta_{fpy} = \beta^w w_{py} + \beta^h_f h_{py} + \pi_{fy} + \mu_p + \xi_{fpy}$ and $\chi_{iop} = \beta^d lnDistance_{op} + \beta^r ruralHukou_i \times \mathbf{1}\{o \neq p\} + \beta^m male_i \times \mathbf{1}\{o \neq p\}$.

- Housing demand equals housing supply:

In the housing market, demand is derived from workers maximizing their Cobb-Douglas preferences for housing and the national good, subject to a budget constraint, as outlined in Section 8.1: $D_{py} = \frac{\zeta W_{py}}{rH_{py}}$. Combined with the housing supply function equation (11), it can be derived that:

$$h_{py} = \Gamma lnL_{py} + \Gamma w_{py} + \varepsilon_{py} \tag{13}$$

where $\Gamma = \frac{\gamma}{1+\gamma}$ and $\varepsilon_{py} = \Gamma(ln\zeta - lnr) + (1-\Gamma)ln\kappa_{py}$

9.2 Policy Impacts on Intergenerational Persistence and Agricultural Employment

For this simple setup, I only need the estimates from the discrete choice model. The key parameter for the housing market, Γ, is calibrated based on the existing literature.[28] I use iteration to obtain counterfactuals. The detailed procedure is described in Appendix Section E.

Using the full model, I estimate the impacts of the same counterfactual policies discussed in Section 8.3 on migration, with results presented in Appendix Table B.12. The patterns are the same as those obtained from the discrete choice model, but the magnitudes of changes in migration behaviors are smaller. This is because housing prices now adjust in response to population changes, which attenuates migration responses. Notably, with a targeted 10% rent subsidy for migrants with low-education fathers, a small number of workers with high-education fathers are displaced as the increased influx of disadvantaged migrants causes a slight rise in Tier-1 housing prices.

Moreover, I offer back-of-the-envelope calculations of the impacts on intergenerational persistence and agricultural employment..

I measure intergenerational persistence as the income gap between workers from different parental backgrounds. On average, workers with high-education fathers earn 22% more than those with low-education fathers. A 10% rent subsidy for all migrants in Tier-1 prefectures increases this gap by 0.59 percentage points, or approximately 3% of the initial difference, as more high-education workers migrate and earn higher incomes than their low-education counterparts. In contrast, a targeted 10% rent subsidy for migrants with low-education fathers reduces the gap by 13%, significantly improving intergenerational mobility. A 10% rent subsidy in non-Tier-1 prefectures narrows the gap by 4%, a smaller effect due to lower income levels in these areas and the simultaneously increased migration among workers with high-education fathers.

For the calculations on agricultural employment, I restrict the sample to workers from rural areas. Those who remain in their origin prefecture are assumed to work in agriculture with a probability equal to the share of rural workers employed in the

[28] The results are similar across different values of Γ in the literature.

agricultural sector within that prefecture. Migrants are assumed to work in the non-agricultural sector. Back-of-the-envelope calculations show that a 10% rent subsidy for all migrants in Tier-1 prefectures increases the agricultural employment gap between workers with low- and high-education fathers by 3%. In contrast, a targeted 10% rent subsidy for migrants with low-education fathers reduces the gap by 6%. A 10% rent subsidy in non-Tier-1 prefectures narrows the gap by 5%. The results imply that to promote structural transformation and encourage rural-to-urban migration, the second and third policies are similarly effective.

10 Conclusion

This paper examines the relationship between housing affordability, internal migration, and intergenerational mobility in China, highlighting the multifaceted impact of rising housing costs on socioeconomic disparities. Utilizing the Housing Purchase Restriction policy as a natural experiment and an instrumental variable approach, this paper documents that elevated housing costs significantly deter internal migration, with disproportionately adverse effects on children from less privileged backgrounds. This dynamic exacerbates income disparities and diminishes intergenerational mobility, reinforcing the cycle of economic disadvantage.

The findings in this paper suggest that in China, a country characterized by geographical inequality (Xie and Zhou, 2014) and low intergenerational mobility, implementing housing price regulations has the potential to alleviate both concerns by encouraging greater internal migration. This implication somewhat relates to the literature explaining the mechanism behind the Great Gatsby Curve, which refers to the positive correlation between inequality and intergenerational persistence observed in many countries. While the literature focuses on education attainment as a main mechanism, this paper points out that even with the same education, access to opportunities is different with different parental backgrounds.

This paper's conceptual framework is relevant in many developing countries where migration plays crucial roles in social mobility. Future research could explore comparative studies across countries with diverse housing markets and migration policies to provide a broader perspective on the global impact of housing affordability on intergenerational mobility.

In conclusion, this research emphasizes the need for targeted policies to address housing affordability and enhance economic equality and social mobility. As housing prices rise globally, it is crucial to understand their impact on migration patterns. Moreover, the findings suggest that high-skill children from disadvantaged families can be trapped in rural areas while low-skill children from affluent families easily migrate. These findings are relevant for the broader discussions on spatial labor misallocation and productivity (Young, 2013). Reducing housing costs could help narrow the gap in migration behaviors, leading to substantial gains in both equity and efficiency.

References

Alberto Alesina, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou. Intergenerational mobility in africa. *Econometrica*, 89(1):1–35, 2021.

Manuela Angelucci. Migration and financial constraints: Evidence from mexico. *Review of Economics and Statistics*, 97(1):224–228, 2015.

Samuel Bazzi. Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, 9(2):219–255, 2017.

Barbara Biasi. School finance equalization increases intergenerational mobility. *Journal of Labor Economics*, 41(1):1–38, 2023.

George J Borjas. Self-selection and the earnings of immigrants. Technical report, National Bureau of Economic Research, 1987.

Kirill Borusyak and Peter Hull. Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185, 2023.

Kirill Borusyak, Peter Hull, and Xavier Jaravel. Quasi-experimental shift-share research designs. *The Review of Economic Studies*, 89(1):181–213, 2022.

Loren Brandt and Carsten A Holz. Spatial price differences in china: Estimates and implications. *Economic development and cultural change*, 55(1):43–86, 2006.

Gharad Bryan, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, 82(5):1671–1748, 2014.

Shu Cai. Migration under liquidity constraints: Evidence from randomized credit access in china. *Journal of Development Economics*, 142:102247, 2020.

Ting Chen, Laura Liu, Wei Xiong, Li-An Zhou, et al. Real estate boom and misallocation of capital in china. *Work. Pap., Princeton Univ., Princeton, NJ*, 9, 2017.

Yuanyuan Chen, Le Wang, and Min Zhang. Informal search, bad search?: the effects of job search method on wages among rural migrants in urban china. *Journal of Population Economics*, 31:837–876, 2018.

Yunjie Chen, Yongzheng Liu, and Hailun Wei. The unintended consequences of home-buying restriction in china. *Available at SSRN 4637419*, 2023.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133 (3):1107–1162, 2018.

Daniel Chiquiar and Gordon H Hanson. International migration, self-selection, and the distribution of wages: Evidence from mexico and the united states. *Journal of political Economy*, 113(2):239–281, 2005.

Eric Chyn, Kareem Haggag, and Bryan A Stuart. The effects of racial segregation on intergenerational mobility: Evidence from historical railroad placement. Technical report, National Bureau of Economic Research, 2022.

Rebecca Diamond. The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *American Economic Review*, 106(3):479–524, 2016.

Dave Donaldson and Richard Hornbeck. Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858, 2016.

Zaichao Du and Lin Zhang. Home-purchase restriction, property tax and housing price in china: A counterfactual analysis. *Journal of Econometrics*, 188(2):558–568, 2015.

Yi Fan, Junjian Yi, and Junsen Zhang. Rising intergenerational income persistence in china. *American Economic Journal: Economic Policy*, 13(1):202–230, 2021.

Hanming Fang, Quanlin Gu, Wei Xiong, and Li-An Zhou. Demystifying the chinese housing boom. *NBER macroeconomics annual*, 30(1):105–166, 2016.

James J Feigenbaum. Intergenerational mobility during the great depression. 2015.

Benjamin Feigenberg, Ben Ost, and Javaeria A Qureshi. Omitted variable bias in interacted models: A cautionary tale. *Review of Economics and Statistics*, pages 1–47, 2023.

Markus Frölich. Parametric and nonparametric regression in the presence of endogenous control variables. *International Statistical Review*, 76(2):214–227, 2008.

Li Gan, Zhichao Yin, Nan Jia, Shu Xu, Shuang Ma, Lu Zheng, et al. Data you need to know about china. *Springer Berlin Heidelberg. https://doi*, 10(978):3, 2014.

Peter Ganong and Daniel Shoag. Why has regional income convergence in the us declined? *Journal of Urban Economics*, 102:76–90, 2017.

Carlos Garriga, Aaron Hedlund, Yang Tang, and Ping Wang. Rural-urban migration, structural transformation, and housing markets in china. *American Economic Journal: Macroeconomics*, 15(2):413–440, 2023.

Garance Genicot, Debraj Ray, and Carolina Concha-Arriagada. Upward mobility in developing countries. Technical report, National Bureau of Economic Research, 2024.

Honge Gong, Andrew Leigh, and Xin Meng. Intergenerational income mobility in urban china. *Review of Income and Wealth*, 58(3):481–503, 2012.

Joseph Gyourko, Yang Shen, Jing Wu, and Rongjie Zhang. Land finance in china: Analysis and review. *China Economic Review*, 76:101868, 2022.

Gaurav Khanna, Wenquan Liang, Ahmed Mushfiq Mobarak, and Ran Song. The productivity consequences of pollution-induced migration in china. Technical report, National Bureau of Economic Research, 2021.

David Lagakos, Samuel Marshall, Ahmed Mushfiq Mobarak, Corey Vernot, and Michael E Waugh. Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*, 113:138–154, 2020.

Sang Yoon Tim Lee, Nicolas Roys, and Ananth Seshadri. The causal effect of parents' education on children's earnings. *Manuscript. https://www. ssc. wisc. edu/aseshadr/WorkingPapers/spillover. pdf [accessed 26 October 2021]*, 2024.

Victor Jing Li, Andy Wui Wing Cheng, and Tsun Se Cheong. Home purchase restriction and housing price: A distribution dynamics analysis. *Regional Science and Urban Economics*, 67:1–10, 2017.

Hong Liu, Lili Liu, and Fei Wang. Housing wealth and fertility: evidence from china. *Journal of Population Economics*, 36(1):359–395, 2023.

Darshini Mahadevia, Zhiyan Liu, and Xiuming Yuan. Housing of urban migrants in india and china. *Environment and Urbanization Asia*, 3(1):143–163, 2012.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, 1973.

Melanie Morten and Jaqueline Oliveira. The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, 16(2):389–421, 2024.

Kaivan Munshi and Mark Rosenzweig. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(01):46–98, 2016.

Martin Nybom and Jan Stuhler. Biases in standard measures of intergenerational income dependence. *Journal of Human Resources*, 52(3):800–825, 2017.

John Parman. American mobility and the expansion of public education. *The Journal of Economic History*, 71(1):105–132, 2011.

Kenneth S Rogoff and Yuanchen Yang. Peak china housing. Technical report, National Bureau of Economic Research, 2020.

Albert Saiz. The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3):1253–1296, 2010.

Gary Solon. Intergenerational income mobility in the united states. *The American Economic Review*, pages 393–408, 1992.

James H Stock and Mark W Watson. *Introduction to econometrics*. Pearson, 2020.

Ang Sun and Qinghua Zhang. Who marries whom in a surging housing market?. *Journal of Development Economics*, 146:102492, 2020.

Xin Wang and Yi Wen. Housing prices and the high chinese saving rate puzzle. *China Economic Review*, 23(2):265–283, 2012.

Zachary Ward. Internal migration, education, and intergenerational mobility: Evidence from american history. *Journal of Human Resources*, 57(6):1981–2011, 2022.

Yu Xie and Xiang Zhou. Income inequality in today's china. *Proceedings of the national academy of Sciences*, 111(19):6928–6933, 2014.

Alwyn Young. Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics*, 128(4):1727–1785, 2013.

Renjie Zhao and Jiakai Zhang. Rent-tax substitution and its impact on firms: Evidence from housing purchase limits policy in china. *Regional Science and Urban Economics*, 96:103804, 2022.

Yaohui Zhao. Leaving the countryside: rural-to-urban migration decisions in china. *American Economic Review*, 89(2):281–286, 1999.

A Appendix: Continuous Model

This section shows a continuous version of the model where, instead of High and Low type, both children's income and fathers' socioeconomic status are continuous. The model is based on Borjas (1987) and Chiquiar and Hanson (2005). The model suggests that housing price changes may increase or decrease migration and then intergenerational mobility, depending on parameters and specific distributions of disutility with respect to migration.

I assume that if a child stays at the origin, they get $Y_{stay} = \mu_0 + \delta_0 F$; if the children migrate, they get $Y_{migrate} = \mu_1 + \delta_1 F$. Migration requires a cost of $M(h, F)$, which depends on the housing price gap h and the father's socioeconomic status F. I assume that $\frac{\partial M(h,F)}{\partial h} > 0$, meaning that an increase in housing price gaps increases migration cost. People also have mental costs for migration, which I call "disutility" and denote as τ. τ is independent of F and the realized Y. I denote the CDF and PDF function of τ as $F_\tau(\tau)$ and $f_\tau(\tau)$

This model makes the following assumptions. First, instead of analyzing the heterogeneity with respect to the potential migrants' education as in Borjas (1987), I assume here that each individual has the same level of education, and the father's socioeconomic status, F, affects the children's income, Y. Second, the disutility function is independent of F and the realized Y.

Combining the above benefit and cost, a child will migrate if:

$$Y_{migrate} - M(h, F) - \tau > Y_{stay} \tag{14}$$

Rearrange the equation and substitute $Y_{migrate}$ and Y_{stay}, the above equation is equivalent to:

$$(\mu_1 - \mu_0) + (\delta_1 - \delta_0)F - M(h, F) > \tau \tag{15}$$

Denote the left-hand side as $B(h, F)$, which is the net benefit of migrating. For a child faced with their father's socioeconomic status F and housing price gap h, the probability that they migrate is $Pr(migrate|h, F) = Pr(B(h, F) > \tau) = F_\tau(B(h, F))$.

In my main regression analysis, I regress the migration indicator on the interaction of paternal socioeconomic status and the housing price gap. The coefficient of interest, the cross derivative of migration probability with respect to the housing price gap and the father's socioeconomic status, can be derived from the model.

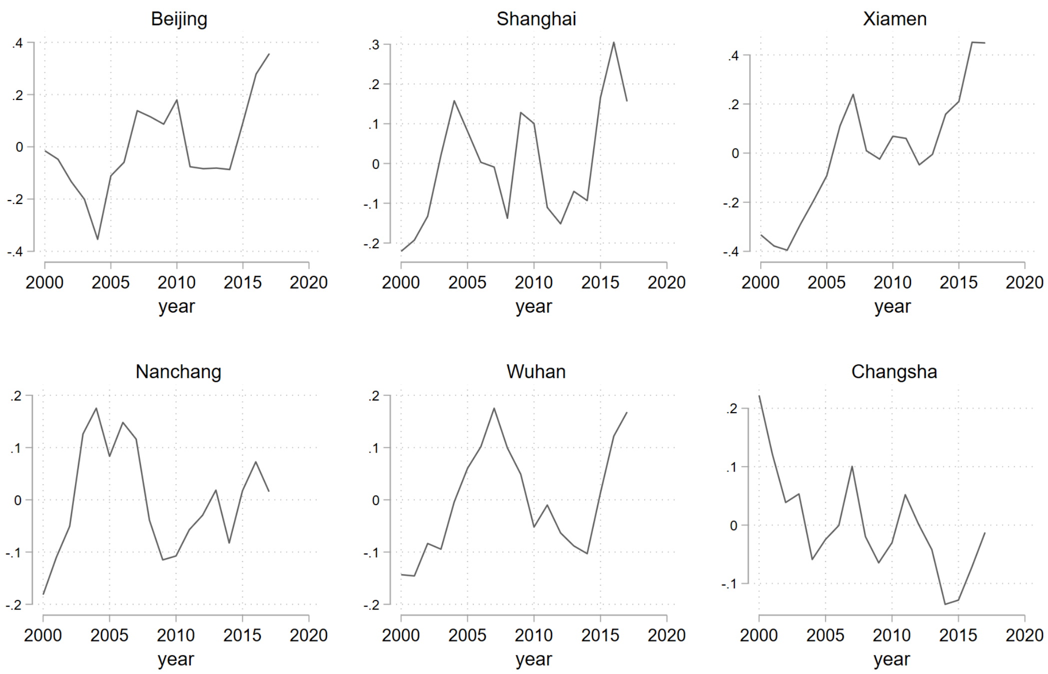
$$\frac{\partial^2 Pr(migrate|h,F)}{\partial h \partial F} = \underbrace{[-f_\tau(B)M''_{Fh}]}_{\text{heterogeneous cost}} + \underbrace{[-f'_\tau(B)B'_F M'_h]}_{\text{position at disutlity distribution}}$$

Similarly, I derive the cross derivative of the expected income with respect to h and F:

$$\frac{\partial^2 E[Y|h,F]}{\partial h \partial F} = \underbrace{[-(B+M)f_\tau(B)M''_{Fh}]}_{\text{heterogeneous cost}} + \underbrace{[-(B+M)f'_\tau(B)B'_F M'_h]}_{\text{position at disutility distribution}} + \underbrace{[-(\delta_1-\delta_0)f_\tau(B)M'_h]}_{\text{heterogeneous benefit}}$$

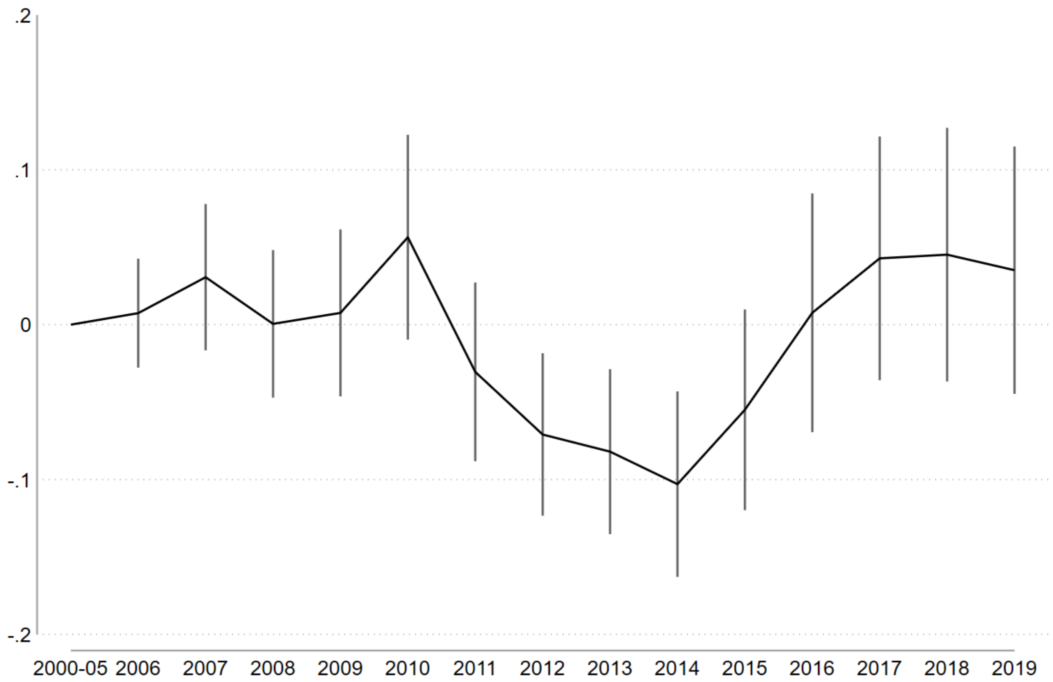
B Appendix: Figures and Tables

Figure B.1: Housing price residuals in six major prefectures in China



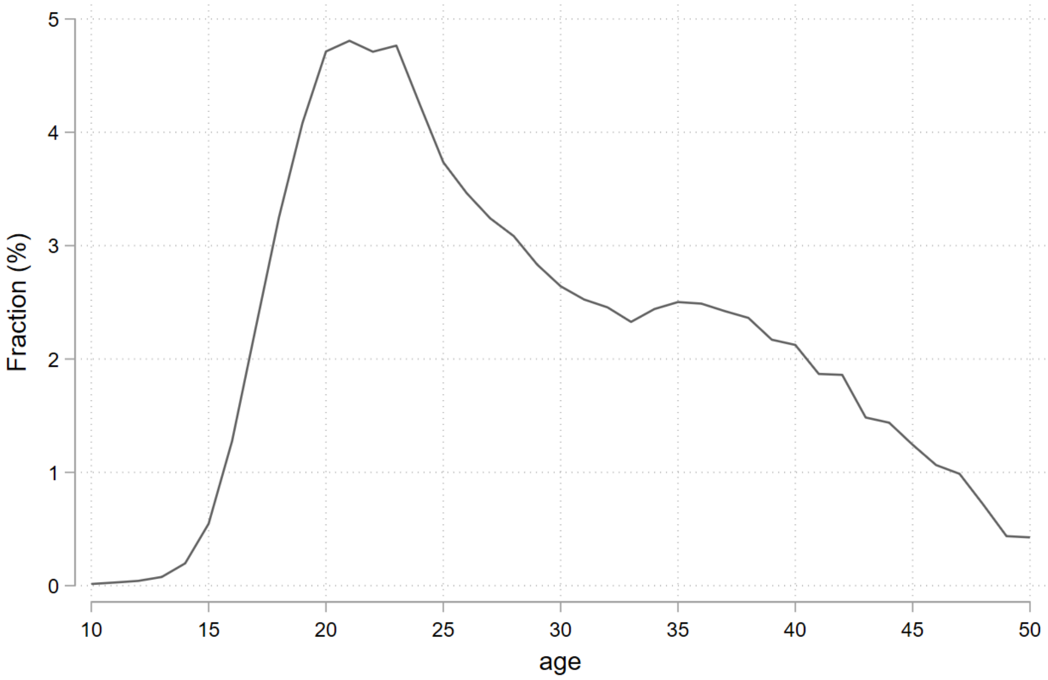
NOTE: The figure shows the residualized housing prices obtained from regressing housing prices (measured by log yuan per square meter in 2000RMB) on prefecture fixed effects and year fixed effects. The figure underscores the considerable temporal and regional variation across prefectures.

Figure B.2: HPR Event Study Plot



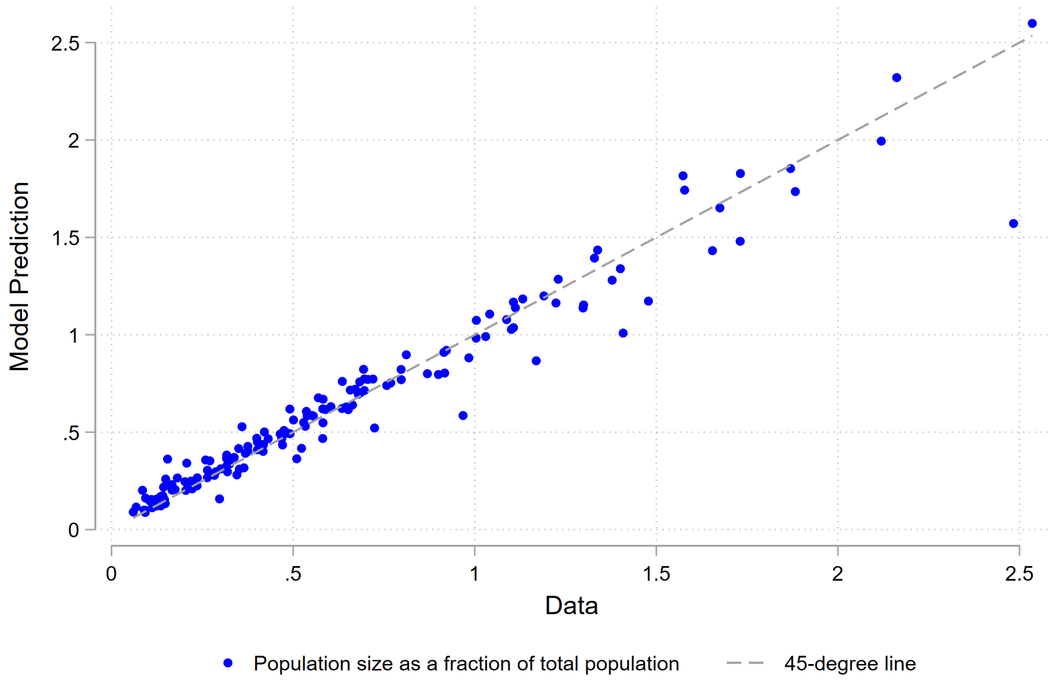
NOTE: The figure shows the event study plot for the HPR policy. Using the prefecture-year-level panel data, housing prices are regressed on prefecture fixed effects, year fixed effects, and interaction terms between year dummies and an indicator for ever adopted the policy. The figure shows the estimates and confidence intervals for the coefficients of the interaction terms, which reflect the changes in differences in housing prices between the ever-treated and never-treated prefectures over time.

Figure B.3: Age-specific Migration Probability



NOTE: The figure shows the age-specific migration probability rate, which is the proportion of people who migrated at a specific age out of the total number of migrants.

Figure B.4: Goodness of Fit: Predicted vs. Actual Population Size



NOTE: The figure shows the observed and predicted population sizes for each prefecture as a fraction of the total population. Each dot represents a specific prefecture.

Table B.1: The List of 46 Prefectures that Adopted the Housing Purchase Restriction Policy with Dates of Announcing and Abolishing the Policy

Prefecture	Start Year	Start Month	End Year	End Month
Beijing	2010	4		
Tianjin	2010	10	2014	10
Shijiazhuang	2011	2	2014	9
Taiyuan	2011	1	2014	8
Huhehaote	2011	4	2014	6
Shenyang	2011	3	2014	9
Dalian	2011	3	2014	9
Changchun	2011	5	2015	6
Haerbin	2011	2	2014	8
Shanghai	2010	10		
Nanjing	2010	10	2014	9
Wuxi	2011	2	2014	8
Xuzhou	2011	5	2014	8
Suzhou	2011	3	2014	9
Hangzhou	2010	10	2014	8
Ningbo	2010	10	2014	7
Wenzhou	2011	3	2013	8
Shaoxing	2011	8	2014	8
Jinhua	2011	3	2014	8
Quzhou	2011	9	2014	7
Zhoushan	2011	8	2013	1
Taizhou	2011	8	2014	8
HeFei	2011	1	2014	8
Fuzhou	2010	10	2014	8
Xiamen	2010	10	2014	8
Nanchang	2011	2	2014	7
Jinan	2011	1	2014	7
Qinghai	2011	1	2014	9
Zhengzhou	2011	1	2014	8
Wuhan	2011	1	2014	7
Changsha	2011	3	2014	8
Guangzhou	2010	10		
Shenzhen	2010	9		
Zhuhai	2011	11	2016	3
Foshan	2011	3	2014	8
Nanning	2011	3	2014	10
Haikou	2010	10	2014	7
Sanya	2010	10		
Chengdu	2011	2	2015	1
Guiyang	2011	2	2014	9
Kunming	2011	1	2014	8
Xian	2011	3	2014	9
Lanzhou	2011	3	2014	7
Xining	2011	8	2014	9
Yinchuan	2011	2	2014	8
Wulumuqi	2011	3	2014	8

NOTE: This table lists the dates of announcing and abolishing the Housing Purchase Restriction (HPR) Policy according to Chen et al. (2017).

Table B.2: Descriptive Statistics

	Mean	SD
Migration	0.20	0.40
lnIncome	9.80	1.67
FatherEduy	5.74	4.51
HPgap	0.51	0.51
Male	0.46	0.50
Eduy	10.05	3.75
Rural Hukou	0.67	0.47
Age	43.19	8.95

NOTE: This table presents descriptive statistics for the main sample, which includes 14,976 observations. While the table displays mean values, fathers' years of education and the housing price gap have been demeaned in the analyses.

Table B.3: IV Results: First Stage

	(1) HPgap	(2) FatherEduy× HPgap	(3) HPgap	(4) FatherEduy× HPgap
HPR	-0.161***	3.049***	-0.163***	3.166***
	(0.014)	(0.438)	(0.014)	(0.434)
FatherEduy× HPR	0.000	2.647***	-0.000	2.709***
	(0.001)	(0.057)	(0.001)	(0.057)
FatherEduy× birthyearFE	Y	Y	Y	Y
Eduy× birthyearFE	-	-	Y	Y
Obs.	14,976	14,976	14,976	14,976
Adj. R-sq	0.995	0.494	0.995	0.495

NOTE: The figure shows the first-stage results. When included, Eduy×HPgap is also instrumented using Eduy×HPR. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.4: Robustness: Additional Controls

	(1) Migration	(2) Migration	(3) Migration	(4) Migration	(5) lnIncome	(6) lnIncome	(7) lnIncome	(8) lnIncome
FatherEduy× HPgap	0.010***	0.009***	0.009***	0.006**	0.033***	0.036***	0.022*	0.029**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.011)	(0.012)	(0.012)	(0.012)
HPgap			-0.716**	-0.656**			-0.562	-0.718
			(0.291)	(0.286)			(1.552)	(1.552)
Eduy× HPgap		0.004		0.009***		-0.012		-0.022
		(0.004)		(0.003)		(0.017)		(0.014)
OriginFE× BirthyearFE	Y	Y	-	-	Y	Y	-	-
Interaction terms	-	-	Y	Y	-	-	Y	Y
Obs.	14,976	14,976	14,976	14,976	12,068	12,068	12,068	12,068
Mean(Dep. Var.)	0.202	0.202	0.202	0.202	9.799	9.799	9.799	9.799

NOTE: This table shows the results of adding more controls. Fathers' years of education interacted with birth year fixed effects are controlled in all columns. Control variables include gender, parental *Hukou* status, log wage, log employment, and log export at the origin prefecture, the weighted average of the log wage, log employment, and log export at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education.

Table B.5: Robustness: Sample Construction

	(1) Migration	(2) Migration	(3) Migration	(4) Migration	(5) lnIncome	(6) lnIncome	(7) lnIncome	(8) lnIncome
FatherEduy× HPgap	0.009***	0.009***	0.009***	0.009***	0.027**	0.027**	0.027**	0.027**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.013)	(0.013)	(0.013)	(0.013)
HPgap	-0.699*	-0.699*	-0.699*	-0.699*	-3.009*	-3.009*	-3.009*	-3.009*
	(0.371)	(0.371)	(0.371)	(0.371)	(1.600)	(1.600)	(1.600)	(1.600)
Exclude	-	adjacent to tier-1	high GDP per capita	popular destinations	-	adjacent to tier-1	high GDP per capita	popular destinations
Obs.	14,976	14,976	14,976	14,976	12,068	12,068	12,068	12,068
Mean(Dep. Var.)	0.202	0.202	0.202	0.202	9.798	9.798	9.798	9.798

NOTE: This table shows the robustness results for different samples. Fathers' years of education interacted with birth year fixed effects and individual's education interacted with the housing price gap are controlled for in all columns. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and the individual's own education level. For Columns (1)-(4), log wage, log employment, and log export at the origin prefecture, the weighted average of the log wage, log employment, and log export at the potential destination prefectures are controlled in addition. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.6: Robustness: Alternative Housing Price Dataset

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
FatherEduy× HPgap	0.011***	0.008***	0.025**	0.036***
	(0.002)	(0.002)	(0.010)	(0.011)
HPgap	-1.738***	-1.559***	-4.146*	-5.226**
	(0.564)	(0.542)	(2.203)	(2.291)
Eduy× HPgap		0.009***		-0.034**
		(0.003)		(0.015)
K-P F-stat	19.678	13.592	27.980	19.316
Obs.	14,950	14,950	12,049	12,049
Mean(Dep. Var.)	0.201	0.201	9.799	9.799

NOTE: This table shows the results leveraging the Anjuke data using the HPR policy treatment as the instrumental variable. Fathers' years of education interacted with birth year fixed effects are controlled for in all columns. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and the individual's own education level. For Columns (1) and (2), log wage, log employment, and log export at the origin prefecture, the weighted average of the log wage, log employment, and log export at the potential destination prefectures are controlled in addition. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.7: Robustness: Average Income Across Waves

	(1)	(2)	(3)	(4)
	lnIncome	lnIncome	lnIncome multi-year	lnIncome multi-year
FatherEduy× HPgap	0.021**	0.027***	0.022***	0.027***
	(0.009)	(0.010)	(0.008)	(0.009)
HPgap	-2.720**	-3.009**	-3.082**	-3.349***
	(1.327)	(1.351)	(1.240)	(1.256)
Eduy× HPgap		-0.016		-0.014
		(0.013)		(0.012)
K-P F-stat	62.223	41.908	60.816	40.854
LM test	107.637	110.246	105.610	107.910
Obs.	12,068	12,068	12,843	12,843
Mean(Dep. Var.)	9.798	9.798	9.317	9.317

NOTE: This table presents results using the log of average income across waves, as opposed to using children's income solely from 2017. The average income calculation only includes data from waves where the worker's residential prefecture remained the same as that of 2017. Notably, about 66% of the observations incorporate income data from multiple waves. For these observations, the income is averaged across the relevant waves and adjusted for the Consumer Price Index (CPI). The analysis controls for fixed effects associated with the observed waves, such as 'observed in both 2015 and 2017'. Additionally, all columns include controls for fathers' education interacted with birth year fixed effects. Columns (1) and (2) replicate results from Table 3 Columns (5) and (6). Differences in observations between Columns (3) and (4) versus (1) and (2) arise when income data from 2015 or 2019 are available for workers missing income data in 2017. Standard errors are clustered at the prefecture-birth year level, with significance levels indicated as follows: ***p<0.01, **p<0.05, *p<0.1.

Table B.8: Robustness: Residualized Fathers' Education

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
FatherEduy_r× HPgap	0.009***	0.008***	0.024**	0.027**
	(0.002)	(0.002)	(0.012)	(0.012)
HPgap	-0.818***	-0.818***	-2.493*	-2.474*
	(0.315)	(0.315)	(1.392)	(1.392)
Eduy× HPgap		0.002		-0.014
		(0.004)		(0.017)
K-P F-stat	35.520	23.678	53.252	35.505
LM test	61.091	61.091	91.183	91.211
Obs.	13,046	13,046	10,490	10,490
Mean(Dep. Var.)	0.147	0.147	9.743	9.743

NOTE: This table shows the results using residualized fathers' years of education to measure parental background. The number of observations differs from the main analyses because some prefecture-by-birth-year cells contain singletons. These singletons are dropped because the residualized father's education cannot be obtained for them. Residualized fathers' years of education interacted with birth year fixed effects are controlled for in all columns. In Columns (1) and (3), the endogenous variables are $HPgap$ and $FatherEduy\_r \times HPgap$; the instrumental variables are HPR and $FatherEduy\_r \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEduy\_r \times HPgap$, and $Eduy \times HPgap$; the instrumental variables are HPR, $FatherEduy\_r \times HPR$, and $Eduy \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted averages of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.9: Intergenerational Income Elasticity (IGE)

	(1)	(2)
	lnIncome	lnIncome
lnFatherIncome	0.381***	0.384***
	(0.059)	(0.059)
lnFatherIncome× HPgap	0.311*	0.392**
	(0.167)	(0.183)
HPgap	-3.107**	-3.373***
	(1.267)	(1.294)
Eduy× HPgap		-0.014
		(0.013)
K-P F-stat	67.002	44.822
LM test	113.223	115.546
Obs.	12,051	12,051
Mean(Dep. Var.)	9.798	9.798

NOTE: This table shows the results using imputed fathers' income in place of fathers' education. In Column (1), the endogenous variables are $HPgap$ and $FatherIncome \times HPgap$; the instrumental variables are HPR and $FatherIncome \times HPR$. In Columns (2), the endogenous variables are $HPgap$, $FatherIncome \times HPgap$, and $Eduy \times HPgap$; the instrumental variables are HPR, $FatherIncome \times HPR$, and $Eduy \times HPR$. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and education. The K-P F-stat row shows the F-stats of the first stages. The LM test row shows the statistics for the underidentification test. A rejection of the null indicates that the matrix of first-stage coefficients has full column rank, i.e., the model is identified. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.10: Robustness: Impute Household Assets

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
lnHHAssets× HPgap	0.128***	0.092***	0.246**	0.301**
	(0.025)	(0.027)	(0.107)	(0.117)
HPgap	-0.904***	-0.796***	-2.827**	-3.105**
	(0.304)	(0.296)	(1.282)	(1.318)
Eduy× HPgap		0.009***		-0.013
		(0.003)		(0.013)
lnHHAssets× birthyearFE	Y	Y	Y	Y
K-P F-stat	37.999	25.726	64.315	43.163
LM test	67.735	68.649	108.986	111.979
Obs.	14,955	14,955	12,051	12,051
Mean(Dep. Var.)	0.202	0.202	9.798	9.798

NOTE: This table shows the results using imputed household assets in place of fathers' education. Imputed household assets interacted with birth year fixed effects are controlled in all columns. In Columns (1) and (3), the endogenous variables are $HPgap$ and $HHAssets \times HPgap$; the instrumental variables are HPR and $HHAssets \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $HHAssets \times HPgap$, and $Eduy \times HPgap$; the instrumental variables are HPR, $HHAssets \times HPR$, and $Eduy \times HPR$. Control variables include gender, parental *Hukou* status, fixed effects for origin prefecture, birth year, and the individual's own education level. For Columns (1) and (2), log wage, log employment, and log export at the origin prefecture, the weighted average of the log wage, log employment, and log export at the potential destination prefectures are controlled in addition. The K-P F-stat row shows the F-stats of the first stages. The LM test row shows the statistics for the underidentification test. A rejection of the null indicates that the matrix of first-stage coefficients has full column rank, i.e., the model is identified. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.11: Conditional Logit: First Stage

	(1) HousingPrices	(2) FatherEduHigh× HP	(3) HousingPrices	(4) FatherEduHigh× HP
HPR	-0.140***	-0.419***	-0.121***	-0.407***
	(0.027)	(0.038)	(0.030)	(0.039)
FatherEduHigh× HPR	0.002	0.697***	0.002	0.692***
	(0.003)	(0.069)	(0.003)	(0.069)
FatherEduHigh× birthcohort FE	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Controls	-	-	Y	Y
Obs.	607	607	603	603

NOTE: Standard errors are clustered at the prefecture level. HP is the abbreviation for housing prices. Columns (1) and (2) show the estimates without macroeconomic controls. Columns (3) and (4) show the results with macroeconomic controls, including employment rate, GDP per capita, log wage, and log export.

Table B.12: Counterfactuals: Spatial Equilibrium Model

FatherEdu	Change in Percentage Points (%) Stay	Migrate to Tier-1	Migrate to Non-Tier-1
Panel A: 10% Rent Subsidy in Tier-1 Prefectures			
L	-0.79	1.01	-0.22
H	-1.20	1.70	-0.49
Panel B: 10% Rent Subsidy in Tier-1 Prefectures for Migrants with Low-Education Fathers			
L	-0.92	1.18	-0.26
H	0.23	-0.29	0.06
Panel C: 10% Rent Subsidy in Non-Tier-1 Prefectures			
L	-3.25	-0.21	3.46
H	-2.35	-0.49	2.84

NOTE: This table shows the percentage point changes in migration behavior among workers from different parental backgrounds for each policy. For example, the -0.79 percentage point change in Panel A indicates the reduction in the proportion of workers with low-education fathers staying in their origin prefecture, a change from 85.15% to 84.36%.

C Appendix: Discussion on Whether to Control for Macroeconomic Conditions

When using migration as the outcome variable, macroeconomic conditions should be controlled for, as they are potential confounders. For example, housing price changes can influence wages through their impact on the construction industry. Since wages affect migration decisions, they are an omitted variable and must be included as a control. The same argument applies to other macroeconomic factors.[29] In practice, including the macroeconomic conditions results in minimal changes, so I do not consider them a major concern.

When using income as the outcome variable, concerns arise regarding the issue of bad control in addition to omitted variable bias. The argument for omitted variable bias is similar to the discussion above. Without controlling for macroeconomic conditions, the impact of HPR on migration reflects the combined effects of changes in housing prices and macroeconomic conditions. Since higher housing prices reducing migration and better macroeconomic conditions increasing it, the estimated effect on migration—and consequently on income—is attenuated.

On the other hand, including macroeconomic controls can lead to bad control issues which also attenuate the estimates. Housing prices affect income by influencing a worker's choice of residential location, and current macroeconomic conditions in that location impact income. Since these conditions are the pathway through which housing prices affect income, controlling for them introduces the bad control problem. Controlling for historical macroeconomic conditions can cause similar issues, as historical and current conditions are highly correlated, potentially absorbing some of the impact of housing prices on income.

Table C.1 presents the results for both outcomes, without and with macroeconomic controls. The coefficients of interest remain similar across specifications, indicating that these controls do not significantly affect the results. In the main analysis, macroeconomic

[29] I acknowledge that wages are endogenous when included in the regression equation, making the coefficient of wage inconsistent. However, this does not affect the consistency of the housing price coefficient, provided the instrumental variable is valid—i.e., uncorrelated with the residual conditional on wage and other control variables. The literature typically assumes that such conditional mean independence of the error term is sufficient for a causal interpretation of coefficients. This interpretation, however, relies on the assumption that the conditional mean of the error term is linear in the control variables, a standard assumption in the literature (Stock and Watson, 2020). If nonlinearity is a significant issue and the relationship between the independent variable of interest and control variables is not linear, the coefficient may be inconsistent (Frölich, 2008). Given that the results are robust to the inclusion of macroeconomic conditions and that nonlinearity concerns are typically assumed away in standard econometric handbooks (Stock and Watson, 2020), I do not consider potential nonlinearity a major threat to the validity of the results.

Table C.1: Robustness: Macroeconomic Control Variables

	(1) Migration	(2) Migration	(3) Migration	(4) Migration	(5) lnIncome	(6) lnIncome	(7) lnIncome	(8) lnIncome
FatherEduy× HPgap	0.011***	0.011***	0.009***	0.009***	0.021**	0.022**	0.027***	0.027***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.009)	(0.009)	(0.010)	(0.010)
HPgap	-1.064***	-0.779**	-0.947***	-0.699**	-2.720**	-2.770*	-3.009**	-2.981**
	(0.289)	(0.303)	(0.280)	(0.297)	(1.327)	(1.510)	(1.351)	(1.518)
Eduy× HPgap			0.009***	0.008***			-0.016	-0.016
			(0.003)	(0.003)			(0.013)	(0.013)
Macroeconomic Controls	-	Y	-	Y	-	Y	-	Y
FatherEduy× birthyearFE	Y	Y	Y	Y	Y	Y	Y	Y
K-P F-stat	43.241	37.685	29.381	25.508	62.223	44.568	41.908	30.220
LM test	80.792	67.772	82.021	68.667	107.637	85.255	110.246	86.392
Obs.	14,976	14,976	14,976	14,976	12,068	12,068	12,068	12,068
Mean(Dep. Var.)	0.202	0.202	0.202	0.202	9.798	9.798	9.798	9.798

NOTE: This table presents robustness results with and without macroeconomic control variables. Even-numbered columns exclude macroeconomic controls, while odd-numbered columns include them. Columns (2), (4), (5), and (7) are the estimates in the main results Table 3. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

controls are included for migration outcomes to address potential omitted variable bias, but excluded for income outcomes to avoid bad control issues.

D Appendix: Procedure of Imputation

The imputation takes the following steps. First, I use CHFS 2015 data, restrict the sample to males, and estimate the following equation:

$$ln(Income)_i = \alpha_0 + \alpha_1 Age_i + \alpha_2 Age\_squared_i + X_i' \alpha_X + \epsilon_i \qquad (16)$$

X is a set of demographic and socioeconomic variables, including fixed effects for education, *Hukou* type, Communist party membership, job position, and birth cohort. There is also a coastal dummy, which accounts for regional differences.[30] Then, based on Equation (16), I calculate the predicted income for my main sample by applying the estimated coefficients to the available information on their fathers' characteristics.[31]

To evaluate the goodness of the imputation, I compare my estimates of the intergenerational income elasticity (IGE) to the existing literature. I show results in Appendix Table B.9 where the $FatherIncome \times birthyear$ fixed effects are dropped so that I observe the estimated coefficient of fathers' log income, i.e., the estimates of IGE. My IGE estimate is approximately 0.384, closely aligning with Fan et al. (2021), who report an IGE of 0.390 for the 1970-1980 birth cohort and 0.442 for the 1981-1988 birth cohort.[32]

[30] Communist party membership is an indicator of whether the father is a member of the Communist Party. Job position refers to administrative levels including 1) ordinary worker, 2) department manager, 3) general manager, 4) (deputy) team leader/section chief, 5) (deputy) division head, 6) (deputy) director, 7) (deputy) bureau head and above, 8) village cadre, 9) township cadre, etc. The coastal dummy equals one if the residential province was a coastal province when the child was age 14. For the birth cohort, I separate fathers into 10-year age groups based on their birth years. When the father's birth year is not available, I use the average birth year of fathers given the child's birth year, which I generated using the 2013 China Health and Retirement Longitudinal Study (CHARLS) data, a nationally representative dataset focusing on the old population.

[31] I do not use the coefficients of age, age squared, or birth cohort because I am using a cross-sectional dataset to impute fathers' income and want to avoid income variation caused by age differences. Instead of excluding age-related variables, an alternative approach involves constructing the imputed income for a worker at a representative age. This alternative approach involves calculating the imputed income under the assumption that all individuals are at the mean age, effectively adding a constant to all observations. The results will be the same, as the imputed income of fathers is centered (demeaned) before estimating the regression models. I also check the results incorporating age-related variables when imputing income, and they remain robust.

[32] The measurement and sample period in my study differ slightly from Fan et al. (2021). They use imputed parents' income rather than fathers' income and focus on individuals born between 1970-1988, while my sample includes individuals born between 1962-2000.

E Appendix: Solving for the Counterfactuals

I use iteration to obtain counterfactuals. Specifically, from the estimation above, I derive the estimates for parameters $\beta^d, \beta^r, \beta^w, \beta^h_{fedu\_L}, \beta^h_{fedu\_H}$ and fixed effects $\{\mu_p\}$ and $\{\pi_{fedu,t}\}$. $\{\xi^{fedu}_{py}\}$ and $\{\varepsilon_{py}\}$ for all the prefectures and birth cohorts are calculated using the following equations:

$$\xi^{fedu}_{py} = \delta^{fedu}_{py} - (\beta^w w_{py} + \beta^h_{fedu\_L} h_{py} + \beta^h_{fedu\_H} h_{py} + \mu_p + \pi_{fedu,t})$$

$$\varepsilon_{py} = h_{py} - \Gamma(lnL_{py} + w_{py})$$

With the estimated parameters, fixed effects, and data on worker characteristics and prefectures, the counterfactual equilibrium is obtained through the following steps:

1. I denote the set of labor supply, housing prices, prefecture average attractiveness for all the prefectures and birth cohorts as $\{L^0_{py}\}$, $\{h^0_{py}\}$, and $\{\delta^0_{fpy}\}$.

2. Increase the housing prices of the prefectures of interest, e.g., Beijing, by 10% via increasing the $\varepsilon_{Beijing}$, capturing changes in the housing supply regulations. Alternatively, I could decrease the housing prices faced by children of low-education fathers by 10%, representing a housing substitution policy.

3. Change in $\varepsilon_{Beijing}$ leads to change in housing price $h_{Beijing}$, which then leads to change in Beijing's attractiveness as a residential prefecture $\delta^{fedu}_{Beijing}$, and eventually affect L_p for all the prefectures, as workers change the migration destinations from Beijing to other prefectures. I denote these new labor supplies in different destinations $\{L^1_{py}\}$.

4. Changes in labor supply will then change the housing prices in the housing market according to Equation (13). A decrease in labor supply in Beijing decreases its housing prices and vice versa for prefectures other than Beijing. A new set of housing prices $\{h^1_{py}\}$ can be obtained by inserting $\{L^1_{py}\}$, $\{w_{py}\}$ and $\{\varepsilon_{py}\}$ into Equation (13).

5. Changes in housing prices will then change the attractiveness of the prefectures according to Equation (10). A new set of attractiveness $\{\delta^1_{fpy}\}$ can be obtained by inserting $\{h^1_{py}\}$, $\{w_{py}\}$, $\{\mu_{py}\}$, $\{\pi_{fedu,t}\}$, and $\{\xi^{fedu}_{py}\}$ into Equation (10).

6. Changes in attractiveness $\{\delta^1_{fedu,py}\}$ of the destinations will affect labor supply again. A new set of labor supply $\{L^2_{py}\}$ can be obtained by inserting $\{\delta^1_{fedu,py}\}$

into Equation (12).

7. I then compare $\{L_{py}^1\}$ and $\{L_{py}^2\}$. If they are close enough, I stop the loop and claim I have reached a new equilibrium. Otherwise, I keep iterating Steps 4-6 until the two subsequent labor supplies $\{L_{py}^s\}$ and $\{L_{py}^{s+1}\}$ are close enough.