Housing Prices, Internal Migration, and Intergenerational Mobility*

Qingyuan Chai†

August 21, 2024

Abstract

This paper examines how housing affordability affects intergenerational mobility in China by influencing internal migration. Utilizing the Housing Purchase Restriction (HPR) policy as a natural experiment and employing an instrumental variable (IV) approach, I find that rising housing costs deter migration, with a more pronounced effect on children from disadvantaged families. Consequently, these children are less likely to migrate and subsequently earn lower incomes than their counterparts from more affluent families, thereby reducing intergenerational mobility. These findings are consistent with a migration decision model where housing expenses disproportionately burden individuals from less affluent backgrounds, possibly due to credit constraints. The results underscore the significance of parental background in shaping children's access to labor market opportunities, even when educational attainment and economic conditions are held constant. To further explore the policy implications, a spatial equilibrium model is developed and estimated to evaluate the effects of various housing policies.

Keywords: Housing Market, Migration, Inequality, Intergenerational Mobility, Structural Transformation
JEL Codes: J62, R23, R31

*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, Suqin Ge, Adam Guren, Douglas Gollin, Joseph Gyourko, Patrick Kline, Dilip Mookherjee, Yuhei Miyauchi, Sangyoon Park, Shing-Yi Wang, 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

The impacts of migration on intergenerational mobility are multifaceted. On the one hand, migration is considered a key mechanism for breaking intergenerational persistence by providing opportunities for higher incomes. On the other hand, the overall impact of migration on societal intergenerational mobility depends on the varied likelihood of migration across different parental backgrounds. If children from lower-income families are less able to migrate and access labor market opportunities, intergenerational mobility may decline. Conversely, if these children relocate more easily, it could boost their economic prospects, enhancing intergenerational mobility. Consequently, the factors that influence migration decisions are crucial, as they distinctly impact families based on their economic status, thereby shaping intergenerational mobility.

This paper focuses on one such factor: housing prices. Specifically, I examine the role of housing prices in shaping internal migration and intergenerational mobility in China. The results imply that, with elevated housing prices, migration decreases, and the decrease is larger for children from less advantaged families. Consequently, these disadvantaged children earn lower incomes compared to their counterparts from more affluent families, leading to a reduction in intergenerational mobility.

While the literature extensively documents intergenerational persistence, the determinants behind it are not fully understood, especially in developing countries (Genicot et al., 2024). Additionally, most research investigates correlates of mobility, while causal evidence is rare. This paper introduces housing costs as a new determinant and provides causal evidence that an increase in housing prices decreases intergenerational mobility. Moreover, existing literature has highlighted a causal relationship between the areas in which individuals grow up and their intergenerational mobility (e.g. Chetty and Hendren, 2018). This paper identifies a different strategic period in life where targeted interventions could be beneficial: policies encouraging migration in adulthood—after children have reached maturity—could enhance intergenerational mobility. Given the recent surge in housing prices observed in many countries, the framework developed in this study may be applicable to various other contexts.

The impact of housing prices on intergenerational mobility is ex-ante ambiguous. On the one hand, elevated housing prices can reduce intergenerational mobility by disproportionally preventing more disadvantaged children from migrating and earning higher incomes. Specifically, migration requires upfront residence costs, as migrants cannot immediately find a job at the destination when they arrive, and there is a searching period. As a result, housing costs can harm intergenerational mobility by disproportionally affecting those who cannot afford it, possibly due to credit constraints (Eggleston et al., 2018; Gai et al., 2021). With disparate abilities among children to capitalize on labor market opportunities due to their parental backgrounds, even if children receive the same level of education and encounter identical economic conditions, family backgrounds can still significantly affect their labor market outcomes.

An opposite direction is also theoretically possible, where increases in housing prices deter migration more for children with privileged backgrounds. Specifically, because migration is very costly, most children from poor backgrounds cannot afford to migrate. The cost barrier can be so high that marginal changes in migration costs will not influence the decision of children from less affluent families to migrate. Conversely, children from wealthier families might be closer to the decision-making threshold regarding migration, making them more responsive to changes in housing prices.[1]

Overall, the sign of housing prices' impact on intergenerational mobility is theoretically undetermined. I construct a model of migration decisions to clarify the conditions under which housing prices either facilitate or hinder intergenerational mobility.

In this paper, China's real estate market serves as a testing ground for examining the migration channel for real estate shocks to affect intergenerational mobility. China provides a unique setting for several reasons. First, China has experienced rapid housing price appreciations over the past decades (Fang et al., 2016), spotlighting the potential impacts of real estate fluctuations. Moreover, housing affordability has become an essential concern for migrants, who complain that "we have no home where there is work, and there is no work where there is home." Second, the real estate boom experienced by different prefectures is substantially heterogeneous, offering rich variations for analyzing its impacts. In particular, the Housing Purchase Restriction policy adopted by some Chinese prefectures during the boom provides a natural experiment to identify the causal effects of real estate shocks. Finally, China's high return to migration is well-documented in the literature, and migration is one of the main approaches to climbing the economic ladder, especially for workers from rural areas (see, e.g., Lagakos et al., 2020).

The phenomenon of rising housing prices impeding migration is observed in many other countries, making this conceptual framework potentially relevant in broader contexts. For instance, recent years have seen rising housing prices in countries such as India and the United States (Olney and Thompson, 2024; Mahadevia et al., 2012). The impact of housing affordability on migration may vary across different parental backgrounds, thereby affecting intergenerational mobility.

[1] Another possibility is that children from wealthier families expect higher housing quality. As the prices of high-quality houses elevate more, rising housing prices can translate into higher migration costs for children with wealthy parents, potentially discouraging their migration.

The empirical results are in line with the credit constraint channel. We observe that higher housing costs lead to reduced migration, but the reduction is attenuated for individuals with fathers of higher education levels. Furthermore, increases in housing cost barriers lead to lower income, but the adverse impact is again mitigated for those whose fathers are more educated.

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 destination prefectures when the economy is booming. As a result, 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, this study employs a natural experiment facilitated by the Housing Purchase Restriction (HPR) policy introduced in China around 2010. By restricting the number of properties that each household or firm could buy, this policy led to an immediate and sharp decrease in housing demand. Since the policy was launched in only 46 prefectures, we can assess its impact by analyzing outcome variations before and after the policy's implementation across both the affected and unaffected prefectures. This policy has been used in many existing papers as an exogenous shock on housing prices (Chen et al., 2017; Zhao and Zhang, 2022; Liu et al., 2023; Chen et al., 2023). In this paper, the implementation of the HPR policy is used as an instrumental variable for housing prices.

Regarding robustness, the results remain quantitatively similar after accounting for diverse economic conditions of both origin and destination prefectures, controlling for the interaction between children's education levels and housing prices, and excluding prefectures with distinct migrant behaviors or those potentially influenced by housing price spillovers from megacities. Additionally, since fathers' education may not be comparable across prefectures or years, I replicate the analyses using a residualized measure of fathers' years of education, obtained by regressing it on fathers' *Hukou* type and prefecture-year fixed effects. I also examine results where fathers' education is replaced with imputed fathers' income or household wealth. All robustness checks yield results similar to the main analyses.

In terms of heterogeneity, the results emphasize the significance of lower education levels in navigating housing cost barriers. Specifically, the critical factor in overcoming housing costs is whether the father has any schooling at all. As long as a father has at least primary school education, higher levels of education, such as junior high school or beyond, offer few additional advantages in overcoming these barriers.

Further analyses also indicate that the effects are more pronounced for children with rural parents. Since migration is often associated with non-agricultural employment, it is also shown that housing prices have varied impacts on agricultural employment depending on parental backgrounds. This suggests an interaction between housing costs, structural transformation, and intergenerational mobility.

The remainder of the paper proceeds as follows. Section 2 discusses the related literature. Section 3 presents background information. Section 4 lays out a model. Section 5 describes the empirical approach. Section 6 introduces the data. Section 7 presents the main results. Section 8 builds and estimates a spatial equilibrium model for counterfactual analyses (The estimation is still in progress), and Section 9 concludes.

2 Relation to the Literature

This paper contributes to several strands of literature. First, it speaks to the literature that investigates the determinants of intergenerational mobility. While the phenomenon of intergenerational persistence is well-documented, the factors driving it are not fully understood. Previous studies highlight the importance of access to education, local labor market conditions, social capital, and economic shocks (for example, Feigenbaum, 2015; Olivetti and Paserman, 2015; Parman, 2011; Zheng and Graham, 2022; Tan, 2023). This paper suggests that despite equal education and economic conditions, a father's socioeconomic status can still shape their children's ability to seize labor market opportunities, thereby impacting their outcomes. Additionally, existing literature predominantly investigates developed countries; this paper focuses on the developing world, where geographic inequality and migration potentially play crucial roles in shaping intergenerational mobility. To the best of my knowledge, this work is the first to examine the impact of housing prices on 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. Additionally, he finds that migration was far more effective than education at allowing children to escape poverty. This paper complements his work by examining a different setting and focusing on the impact of elevated housing prices, a trend currently witnessed in many countries.

Furthermore, this paper connects with research exploring the effects of moving-inducing policies to reduce inequality. For example, Chetty et al. (2016) examine the impact of the Moving to Opportunity (MTO) experiment, where they emphasize the importance of children being exposed to better environments during childhood through the induced migration of their families.[2] In contrast, this study sheds light on policies that could enhance intergenerational mobility by encouraging migration in adulthood. This distinction highlights different strategic periods in life where targeted interventions (childhood vs. adulthood) could aid in breaking the cycle of poverty and dependency on one's parental socioeconomic status.

Secondly, this study enriches the literature on the consequences of changes in house prices amidst China's rapid economic transformation. Recent studies have discussed the effects of house price growth on various household decisions in China, including consumption and saving, labor supply, adult children's co-residence with parents, marriage, and fertility (see, e.g., Wei and Zhang, 2011; Wang and Wen, 2012; Li and Wu, 2019; Li et al., 2020; Sun and Zhang, 2020; Alm et al., 2022; Liu et al., 2023).

Thirdly, this paper relates to the literature on the driving forces of internal migration in China. Prior research has emphasized the impacts of *Hukou* registration system, social network, the land tenure system, demographic forces, and pollution (see, for example, Zhao, 1999; Whalley and Zhang, 2007; Chen et al., 2018; Kinnan et al., 2018; Tombe and Zhu, 2019; Jin and Zhang, 2023; Khanna et al., 2021; Guo et al., 2024). This study complements the literature about the impacts of housing prices on internal migration (Meng et al., 2023). The paper also connects to the literature on housing prices and migration decisions in general. Particularly, in the context of the United States, Ganong and Shoag (2017) find that the deceleration of low-skilled workers relocating to high-cost housing areas can help to explain geographic wage convergence from 1980 to 2010.[3] This paper reveals that the effects of housing affordability on migration vary according to parental background, which in turn impacts intergenerational mobility.

Finally, this paper engages with the literature on poverty traps (Azariadis and Stachurski, 2005) and, specifically, how credit constraint significantly impedes migration in developing contexts. Existing papers have documented that credit constraint is an important barrier to both internal migration (Bryan et al., 2014; Angelucci, 2015; Bazzi, 2017). Fewer studies have pointed out the impacts of credit constraints for internal migration (Cai, 2020; Gai et al., 2021). While internal migration may cost less than international migration, this paper suggests that the barrier remains substantial and

[2] Along this line, Alesina et al. (2021) find that in the context of Africa, an additional year spent in a high-mobility region between the ages of 5 and 12 significantly increases the likelihood that children of uneducated parents will complete primary school.

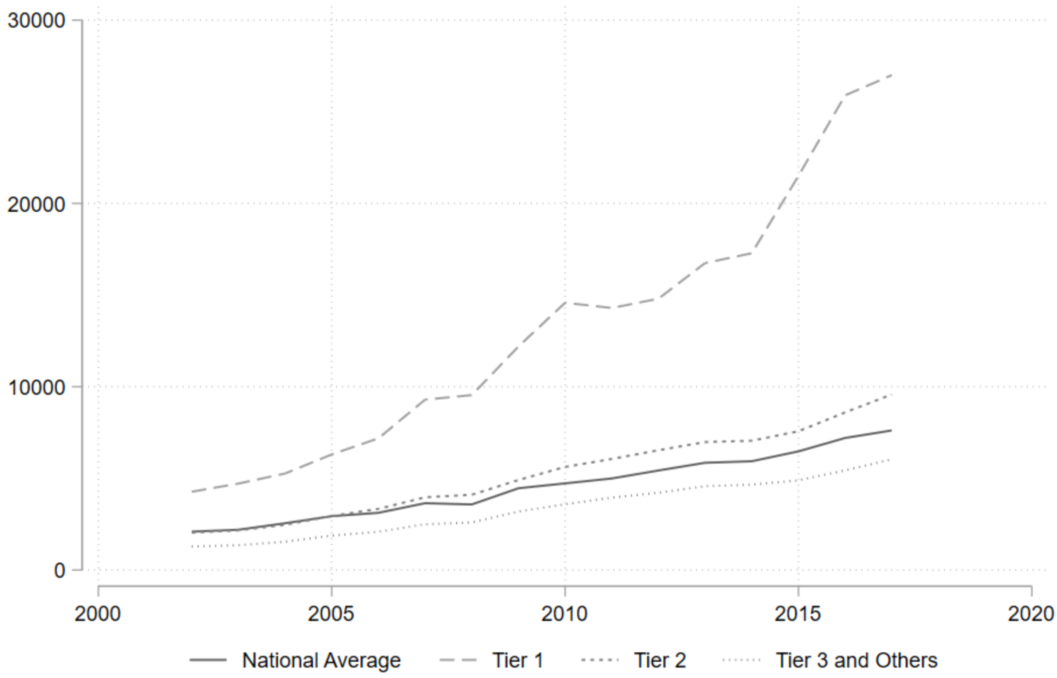
[3] Additionally, some papers explore the heterogeneous impact of housing prices on renters' and owners' migration decisions, which take housing price shocks as income shocks rather than barriers to migration (Zabel, 2012; Foote, 2016). There is a large literature analyzing individual location choices that highlight the role of migration costs in static or partial equilibrium frameworks using quantitative spatial models, as summarized by Redding and Rossi-Hansberg (2017). Closely related to this study, Garriga et al. (2023) investigates the interrelationship between urbanization, structural transformation, and the housing market through the lens of a dynamic spatial equilibrium model.

can affect intergenerational persistence. Given the credit constraint and the fact that migration is risky, a portion of households are stuck in a low-income situation and wary of undertaking the leap to migrate.

3 Background

3.1 Housing Boom in China

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)

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

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.

While there has been a general uptrend in housing prices over the past few decades, significant temporal and regional disparities exist across prefectures. These variations can be attributed to factors such as the differential housing supply elasticity, diverse local house purchasing policies, and spatial inequality in economic development (Chen and Wang, 2015; Saiz, 2010; Gyourko et al., 2022). Figure B.1 illustrates these variations by showing the residualized housing prices, which are derived from regressing housing prices (measured by log yuan per square meter 2000 RMB) on prefecture fixed effects and year fixed effects. The six prefectures are picked randomly from the major prefectures. The figure underscores the considerable temporal and regional variations across prefectures.

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.[4] 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 is 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]The dates of announcing and abolishing the Housing Purchase Restriction (HPR) Policy are from Chen et al. (2017)

4 Model

How do housing prices affect migration decisions and intergenerational mobility? To explore the relevant factors, I constructed 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

Stay

$$
\begin{array}{cc|cc}
 & & \multicolumn{2}{c}{C} \\
 & & H & L \\
\hline
\multirow{2}{*}{F} & H & P_H & 1 - P_H \\
 & L & P_L & 1 - P_L
\end{array}
$$

Migrate

$$
\begin{array}{cc|cc}
 & & \multicolumn{2}{c}{C} \\
 & & H & L \\
\hline
\multirow{2}{*}{F} & H & P'_H & 1 - P'_H \\
 & L & P'_L & 1 - P'_L
\end{array}
$$

In this section, we describe a discrete version of the model for simplicity. A continuous version of the model can be found in the Appendix A where the main results are similar. We assume that both the fathers' and children's incomes are binary, either high (H) or low (L). The probability of the father having a high (H) income is p, which is exogenously given. The value of p does not affect the results of interest. The two matrices above show the potential income for children with fathers of different types. Specifically, we assume that if a child stays at the origin, they have a probability P_F of getting high (H) income. If the child migrates, they have a probability P'_F of getting high (H) income. Both incomes depend on F being high (H) or low (L).

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

The model operates under a set of assumptions. First, it posits that all individuals have the same level of education, and it is the father's type, F, that affects the children's income, C. Second, it is assumed that $P'_H > P_H$ and $P'_L > P_L$; without these conditions, children would not migrate. Third, the model assumes that incomes at the origin and destination are realized after the migration decision. That is, the decision to migrate is made based on the expected income from either staying or migrating rather than on actual income.[5] Lastly, the migration decision is irreversible.

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

$$\begin{aligned} &[P'_F H + (1 - P'_F)L] - [P_F H + (1 - P_F)L] - M(h, F) - \tau > 0 \\ \iff &(P'_F - P_F)(H - L) - M(h, F) > \tau \end{aligned} \tag{1}$$

where F can be 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 our main regression analysis, we regress the indicator for migration on the interaction term of paternal socioeconomic status and housing cost. The coefficient of the interaction term corresponds to $\frac{\partial^2 Pr(migrate|h,F)}{\partial h \partial F}$. As we are analyzing a binary model, the sign of this cross derivative is equivalent to the sign of $(\frac{\partial Pr(migrate|h,F)}{\partial h})_{F=H} - (\frac{\partial Pr(migrate|h,F)}{\partial h})_{F=L}$. It can be derived that:

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

and

$$(\frac{\partial Pr(migrate|h, F)}{\partial h})_{F=H} > (\frac{\partial Pr(migrate|h, F)}{\partial h})_{F=L} \iff \frac{M'_h(H)}{M'_h(L)} < \frac{f_\tau(B_L)}{f_\tau(B_H)} \tag{3}$$

where B_F denotes $B(h, F)$, $F = H$ or L.

In the empirical part, we also analyze the impacts on children's income. Similarly, the sign of the cross derivative of children's expected income with respect to h and F, $\frac{\partial^2 E[C|h,F]}{\partial h \partial F}$, is equivalent to the sign of $(\frac{\partial E[C|h,F]}{\partial h})_{F=H} - (\frac{\partial E[C|h,F]}{\partial h})_{F=L}$. It can be derived that:

$$\frac{\partial E[C|h, F]}{\partial h} = -f_\tau(B(h, F))M'_h(F)(P'_F - P_F)(H - L) < 0 \tag{4}$$

and

$$(\frac{\partial E[C|h, F]}{\partial h})_{F=H} > (\frac{\partial E[C|h, F]}{\partial h})_{F=L} \iff \frac{M'_h(H)}{M'_h(L)} < \frac{f_\tau(B_L)}{f_\tau(B_H)} \frac{(P'_L - P_L)}{(P'_H - P_H)} \tag{5}$$

[5]Mathematical derivations indicate that if the child's income, should they remain, is known before making the migration decision, the expressions differ, but the main insights of the model remain unchanged.

4.2 Discussion

Equation (3) suggests that the impact of housing prices on migration decisions depends on two factors: the heterogeneous costs, $\frac{M'_h(H)}{M'_h(L)}$, and the different position at the disutility distribution, $\frac{f_\tau(B_L)}{f_\tau(B_H)}$. First, we analyze the magnitude of $\frac{M'_h(H)}{M'_h(L)}$. There are at least two potential reasons for $M'_h(F)$ to either increase or decrease in F. If credit constraints exist, housing costs will increase migration costs more for children with L type fathers, namely $M'_h(F)$ will decrease in F. If, on the other hand, children with H type fathers require higher housing quality, $M'_h(F)$ will increase in F, as the same housing price change leads to a larger migration cost increase for children with H type fathers. The overall magnitude of $\frac{M'_h(H)}{M'_h(L)}$ depends on the relative importance of these two potential mechanisms.

For $\frac{f_\tau(B_L)}{f_\tau(B_H)}$, $f_\tau(B_F)$ measures the sensitivity of disutility with respect to the net benefit of migration. If more children of H type fathers are at the margin of deciding whether to migrate or not, while fewer children of L type fathers are at the margin, we have $f_\tau(B_L)$ smaller than $f_\tau(B_H)$, and vice versa.

Equation (5) suggests that the two mechanisms described above, heterogeneous migration costs and differential position at disutility distribution, also apply to the analysis of the heterogeneous impacts of housing costs on children's income. There is an additional term, $\frac{P'_L - P_L}{P'_H - P_H}$, which measures the relative benefit of migrating between children with L and H type fathers.[6]

Overall, the signs of $\frac{\partial^2 Pr(migrate|h,F)}{\partial h \partial F}$ and $\frac{\partial^2 E[C|h,F]}{\partial h \partial F}$ are theoretically undetermined. The empirical analyses help figure out which force dominates in the data.

5 Empirical Approach

We start our analysis by assessing how housing prices impact migration decisions and income across fathers' years of education. Then, to address the potential identification issue, we conduct 2SLS estimation using the Housing Purchase Restriction (HPR) policy to construct the instrumental variable.

[6] If fathers' social network is more valuable in the hometown, $\frac{P'_L - P_L}{P'_H - P_H}$ is larger than one, meaning that migration is more beneficial for children with low-type fathers. Given this assumption, we have Equation (5) as 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.

5.1 Baseline

We run the following regression to examine the heterogeneous impacts of housing prices on migration and income across different parental backgrounds. Specifically, for individual i born in year y whose origin is prefecture p, we estimate the following equation:

$$\begin{aligned} Outcome_{iyp} = &\beta_1 FatherEdu_{iyp} \times HPgap_{yp} + \beta_2 FatherEdu_{iyp} \\ &+ \beta_3 HPgap_{yp} + Z_{yp} + X_{iyp} + \mu_y + \eta_p + \epsilon_{iyp} \end{aligned} \tag{6}$$

where β_1 is the coefficient of interest. $Outcome_{iyp}$ can be either the indicator for migration or log income.[7] Migration is defined as whether the individual lives in a different prefecture from their origin, where the origin is defined as the prefecture that the individual lived in at age 14.[8] Based on the China Migrants Dynamic Monitoring Survey (CMDS) data, over 90% of migrants migrated across prefectures (as compared to within prefecture). $HPgap_{yp}$ is the housing price gap between the origin and potential destinations, which we describe in detail below. $FatherEdu_{iyp}$ is the individual's father's years of education. Control variables include gender, parents' *Hukou* status, employment and wage level at the origin prefecture, weighted averages of employment and wages at the destination prefectures, and fixed effects for origin prefecture, birth year, and education. Standard errors are clustered at prefecture times birth year level.[9]

Although we are using cross-section data from one single year, the survey has information on the migration history of the interviewee, including the origin, destination, and year of migration, allowing us to back out the housing prices that they were faced with during the working ages. To accurately measure the effect of the housing market on migration decisions, ideally, we would like to pair each migration decision with the characteristics of the housing market at the time when the individual was contemplating

[7]One potential concern is that, as there exists a large rural-urban gap in the cost of living, the variations in income may capture mainly the price difference in different areas but not the real difference. To address this concern, we 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. We do not use adjusted income for the main results because, according to anecdotes and as documented in Albert and Monras (2022), a nonnegligible portion of people's consumption occurs at the origin rather than the destination. Moreover, the results change little when using adjusted income.

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

[9]Serial correlation does not seem a big concern in the data. Specifically, following Bertrand et al. (2004), I estimate the first, second, and third autocorrelation coefficients for the mean residuals at prefecture times birth year level from the baseline regression. The autocorrelation coefficients are obtained by a simple OLS regression of the residuals on the corresponding lagged residuals. The estimated first-order autocorrelation coefficient is about -.01. The second- and third-order autocorrelation coefficients are both around .02. None of them is statistically significant. I thus conclude that serial correlation matters little for the estimations and use standard errors clustered at prefecture times birth year level for the results.

whether to migrate. Following Sun and Zhang (2020), the expected housing price gap they were exposed to is calculated as follows.

First, using China's 2010 census sample data, we calculate the age-specific migration probability, $age\_specific\_migration\_prob_k$, which represents the proportion of people who migrated at age k out of the total number of migrants. Appendix figure B.4 plots the age-specific migration probability rate.

$$age\_specific\_migration\_prob_k = \frac{\#\ of\ migrants\ migrated\ at\ age\ k}{\#\ of\ migrants}$$

To minimize the influence of parental migration decisions, we start the calculation at age 16. Additionally, the calculation ends at age 45 because data suggests that people barely migrate after 45. If an individual is of age l in the survey year, they were exposed no later than age l, in which case l is the upper bound of the summation.

Specifically, for an individual born in year y, at age l, originating from prefecture p, the expected housing price gap to which they were exposed is defined by the following expression:[10] [11]

$$HPgap_{yp} = \frac{\sum_{k=16}^{min\{l,45\}} age\_specific\_migration\_prob_k \times (lndestprice_{y+k,p} - lnprice_{y+k,p})}{\sum_{k=16}^{min\{l,45\}} age\_specific\_migration\_prob_k}$$

$HPgap_{yp}$ is a weighted average of the housing price gap between the origin and the potential migration destinations. Specifically, to calculate the housing price of potential destinations, we use the weighted average of housing prices in these destinations, with weights determined by existing social networks. The process begins by using census data to determine the number of migrants between each pair of origin and destination prefectures. The fraction of migrants who migrated to each destination for a specific origin can then be calculated and used as the weights. For example, if 50% of migrants from prefecture p move to Shanghai, then Shanghai's housing price is given a weight of 50% in the calculation for prefecture p. The formula for calculating the weighted average housing price $lndestprice_{pt}$ is as follows:

[10] We assume the age-specific migration rate is homogeneous across prefectures. Results remain robust if we use province-specific migration rate. In addition, we assume the age-specific migration rate does not change over time. To mitigate the concern that the age-specific probability also captures birth cohort effects, we do the same analysis using the probabilities generated leveraging census data in 2000 and 2005. The results are very similar.

[11] For transparency and concerns about the functional form, results are robust using the housing price gap for a certain age, such as 20 or 25.

$$lndestprice_{pt} = \sum_{d} w_{pd} lnprice_{dt}$$

where w_{pd} is the fraction of migrants from origin prefecture p who migrated to prefecture d, and $\sum_d w_{pd} = 1$.[12] $lndestprice_{pt}$ is thus the weighted average of housing price at the potential destinations for origin prefecture p in year t.

Variations in housing price gaps arise from two sources: first, children born in the same prefecture but in different years encounter different housing price shocks at ages crucial for making migration decisions; second, children from different origin prefectures experience disparate housing price shocks given the differing origin and destination prefectures.

The control variables at the prefecture times birth year level undergo a similar process of weighting. Specifically, log wage and log employment at the origin are calculated as weighted averages based on age-specific migration probabilities. For wage and employment at the destinations, the weighted averages are computed in a manner similar to that for housing price gaps, with an initial weighting according to social network factors followed by a secondary weighting using age-specific migration probabilities.

As the housing prices in China persistently increased over the last decades, as shown in Figure 1, there is a concern that the $FatherEdu \times HPgap$ may predominantly reflect heterogeneous impacts of fathers' education across birth cohorts. To address this concern, we further control the interaction between fathers' education and fixed effects for birth years. Specifically, we estimate the following equation:

$$\begin{aligned} Outcome_{iyp} = \beta_1 FatherEdu_{iyp} \times HPgap_{yp} + \gamma^y FatherEdu_{iyp} \\ + \beta_3 HPgap_{yp} + Z_{yp} + X_{iyp} + \mu_y + \eta_p + \epsilon_{iyp} \end{aligned} \tag{7}$$

where, compared to Equation (6), the impact of the fathers' education is allowed to vary for children born in different years. This approach isolates the time-varying effects of fathers' education, thereby refining the estimation of the coefficient of interest, β_1.

[12]For the main results, we use census data in 2010 to generate the social network weights, which has both a large sample size and detailed information on origin and destination prefectures of migrants. While these weights are time-invariant in the regression, it could be concerning that the weights may have included impacts on housing prices. We have tried weights generated using census data in 2005 and the CMDS data in 2010-2017, restricting migrants to those who migrated before 2000. All these results are very similar to the main results.

5.2 Instrumental Variable

Housing prices are highly correlated with local economic conditions. Although we have controlled for various origin and destination economic conditions, unobservables can still drive the results. To address the endogeneity issue, we need an instrumental variable to capture the impacts of housing prices. The Housing Purchase Restriction (HPR) policy is leveraged as a natural experiment to construct the instrumental variable.

Appendix Figure B.2 shows the housing prices time series between the prefectures that ever implemented the HPR policy and those that never adopted the policy, measured by log yuan per square meter adjusted to 2000 RMB. To make the comparison clearer, the housing prices of the never-treated prefectures are shifted upward by 0.88 log points. Visually, the pre-trend was quite parallel before the implementation of the HPR policy around 2010. Since the implementation of the policy, the growth of the treated prefectures immediately got slower, 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.3 shows the event study plot for the policy. Specifically, 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 having adopted the policy. The figure shows the estimations 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. Estimations suggest that the differences remained stable before the policy, became much smaller during the policy, and got back to the original level once the policy was lifted in most of the treated prefectures.

To construct the instrumental variable, we first create a dummy variable at the prefecture-year level, indicating whether the HPR policy is in effect. Then, the dummies for the origin and destination prefectures are weighted and summed in the same way as the housing price gaps. The weighted average, HPR_{yp}, is then used as the instrumental variable for the housing price gap $HPgap_{yp}$.

Specifically, for individual i born in year y whose origin is prefecture p, the first stage equations are:

$$\begin{aligned} FatherEdu_{iyp} \times HPgap_{yp} = {} & \beta_1 FatherEdu_{iyp} \times HPR_{yp} + \beta_2 FatherEdu_{iyp} \\ & + \beta_3 HPR_{yp} + Z_{yp} + X_{iyp} + \mu_y + \eta_p + \epsilon_{iyp} \end{aligned} \tag{8}$$

$$
\begin{aligned}
HPgap_{yp} = \beta_1 FatherEdu_{iyp} \times HPR_{yp} + \beta_2 FatherEdu_{iyp} \\
+ \beta_3 HPR_{yp} + Z_{yp} + X_{iyp} + \mu_y + \eta_p + \epsilon_{iyp}
\end{aligned} \tag{9}
$$

The predicted $\widehat{FatherEdu_{iyp} \times HPgap_{yp}}$ and $\widehat{HPgap_{yp}}$ are then inserted into Equation (6) for second-stage estimation.[13]

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 commercial 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).

It is recognized that, while housing prices are leveraged 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.[14] Therefore, we use housing prices to proxy the residential costs faced by migrants.

The CEIC database also contains data on local economic conditions, including employment, wage, export, FDI, and fiscal expenditure. We complement the dataset by incorporating data on the proportion of employment in the secondary and tertiary sectors sourced from the China City Statistical Yearbook.

6.2 Individual Characteristics

Individual characteristics data are sourced from the China Household Finance Survey (CHFS), a nationally representative survey project executed by the Research Center for China Household Finance. The CHFS aims to gather detailed micro-level data on

[13]As the policy is staggered and on-and-off, the first stage estimations are not consistent. Econometrically, inconsistent estimation of the first stage will not affect the consistency of the second stage estimation as long as the instrumental variable is valid.

[14]A simple OLS regression which regresses log rents on log housing prices has an *F-stats* of 78.22, suggesting a strong correlation between the two stats.

household finance. The data includes rich information on demographic characteristics, employment status, assets and liabilities, income[15] and consumption, social security and insurance, and subjective attitudes. This dataset provides a comprehensive and detailed description of household economic and financial behaviors. The CHFS utilizes a stratified three-stage probability proportion to size (PPS) random sample design, and the data has national and provincial representation.[16] Further information about the dataset can be found in Gan et al. (2014).

This study utilizes data from the 2017 China Household Finance Survey (CHFS), which provides information on the respondents' origin and destination prefectures and migration history.[17] For each individual, the CHFS collects basic information about their parents, such as education, *Hukou* status, and occupations, regardless of whether they co-reside with the interviewee or whether the parents are alive. This feature of CHFS helps us avoid the potential selection problems commonly seen in studies that use surveys that focus on cohabitating household members.

We restrict the sample in the following ways. Firstly, because the migration behavior is potentially different in relatively affluent prefectures compared to the others, we drop observations whose origin is one of the Tier-1 prefectures: Beijing, Shanghai, Shenzhen, and Guangzhou. Secondly, for the availability of income information and to mitigate issues caused by return migration, the observations are restricted to the population aged 16-55.[18]

7 Results

7.1 OLS Results

We estimate Equation (6) using OLS, and the results are shown in Column (1) of Table 1. The coefficient of $HPgap$ is negative, suggesting that increases in housing costs decrease migration. As fathers' education is not centered, the magnitude of the coefficient reflects the impacts of housing prices for children with fathers of zero years of schooling. The coefficient of $FatherEdu$ is positive, indicating that fathers' education can mitigate the

[15]In particular, the survey includes detailed information on income from the informal sector and farming.

[16]Source: https://www.cnopendata.com/en/data/m/chfs/chfs.html

[17]Literature suggests that estimating intergenerational mobility using single-year data can lead to attenuation bias coming from transitory fluctuation in income. We recognize this weakness of our main datasets. When using multiple waves of data, the results remain robust, and the magnitudes of the main coefficients of interest are similar. We stick to single-year data in the main analysis as the other waves all miss some critical information and thus need to merge the information from another wave. Such merging can potentially lead to attribution bias.

[18]The results are not sensitive to different selections of the upper and lower bound of the age range.

impact of housing costs. Despite the influence of fathers' education, it does not fully offset the challenge posed by housing prices. Eliminating this challenge would require more than 30 years of education, which is beyond what is achievable in this setting.

Table 1: Impacts on Migration Decision

	(1) Migration	(2) Migration	(3) Migration	(4) Migration
FatherEdu	0.005***		0.005***	
	(0.001)		(0.001)	
FatherEdu× HPgap	0.008***	0.008***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
HPgap	-0.362***	-0.396***	-0.334***	-0.368***
	(0.078)	(0.079)	(0.078)	(0.078)
Edu× HPgap			0.011***	0.011***
			(0.002)	(0.002)
FatherEdu× birthyearFE	-	Y	-	Y
Obs.	14,050	14,050	14,050	14,050
Adj. R-sq	0.188	0.190	0.191	0.192
Mean(Dep. Var.)	0.140	0.140	0.140	0.140
Mean(Dep. Var.) if FatherEdu=0	0.083	0.083	0.083	0.083

NOTE: This table shows the OLS results. Compared to the odd columns, the even columns control for the fathers' years of education interacted with birth year fixed effects, so the level term of fathers' education year is absorbed, and the estimation is not shown in the table. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

To address the concern that the $FatherEdu \times HPgap$ may predominantly reflect heterogeneous impacts of fathers' education across birth cohorts, we estimate Equation (7) where the effect of the fathers' education is allowed to vary for children born in different years, thus isolating the time-varying impact of a father's education. The results are shown in Column (2) of Table 1. The coefficient of interest barely changes.

Because fathers' education also affects children's education, one potential explanation for the findings is that fathers' education proxies children's education, and it is children's higher education attainment that is overcoming the housing costs. While such a channel also affects intergenerational mobility, the policy implications are different in terms of whether to increase education or labor market opportunities for workers.

To check how much of the observed impact is from children's education, we control for $Edu_{iyp} \times HPgap_{yp}$ in columns (3) and (4). The coefficients of $Edu_{iyp} \times HPgap_{yp}$ are positive and significant, suggesting that children's education attainment helps overcome the housing costs barrier. The coefficients of $FatherEdu_{iyp} \times HPgap_{yp}$ decrease slightly but remain large and significant, suggesting that although the impact of fathers' education on children's education partly explains how fathers' education helps overcome barriers related to housing costs, fathers' education continues to help through channels other than influencing children's education.

Table 2: Impact of housing price on intergenerational mobility

	(1)	(2)	(3)	(4)
	lnIncome	lnIncome	lnIncome	lnIncome
FatherEdu	0.022***		0.022***	
	(0.004)		(0.004)	
FatherEdu× HPgap	0.029***	0.028***	0.020***	0.020***
	(0.007)	(0.007)	(0.007)	(0.007)
HPgap	-0.860**	-0.863**	-0.797**	-0.802**
	(0.373)	(0.383)	(0.373)	(0.383)
Edu× HPgap			0.030***	0.030***
			(0.009)	(0.009)
FatherEdu× birthyearFE	-	Y	-	Y
Obs.	11,308	11,308	11,308	11,308
Adj. R-sq	0.348	0.349	0.349	0.349
Mean(Dep. Var.)	9.736	9.736	9.736	9.736
Mean(Dep. Var.) if FatherEdu=0	9.123	9.123	9.123	9.123

NOTE: This table shows the OLS results. Compared to the odd columns, the even columns control for the father's year education interacted with birth year fixed effects, so the level term of fathers' education year is absorbed, and the estimation is not shown in the table. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table 2 shows the results estimated in the same way as in Table 1 with the outcome replaced with the log income of the children. The sample sizes shrink as only observations with available income information are included. The coefficient of $FatherEdu$ is positive and significant, suggesting that fathers' education is positively correlated with the children's income. 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 percent on average. Our estimation is somewhat larger.

The negative coefficients of $HPgap$ suggest that higher housing costs decrease income. This aligns with our expectations, as migration typically increases income, while higher $HPgap$ values reduce migration. As $FatherEdu$ is not centered, the magnitude of the coefficient reflects the impacts of housing prices for children with fathers of zero years of education. The coefficients of $FatherEdu \times HPgap$ are positive, indicating that fathers' education can mitigate the impact of housing costs.

Columns (2) and (4) control for fathers' years of education interacted with birth year fixed effects, leading to little change in the estimations. Columns (3) and (4) control for $Edu \times HPgap$, the coefficients of interest decrease slightly but remain large, positive, and statistically significant.

7.2 IV Results

The OLS results are subject to omitted variable bias. Specifically, unobserved factors, such as job opportunities, might confound housing prices. As a booming economy attracts migrants while higher housing costs deter migration, we expect the coefficient of $HPgap$ to be underestimated in the OLS results. The coefficient of $FatherEdu \times HPgap$ may indicate that fathers with higher education levels are more effective at leveraging their social networks to secure job opportunities for their children in destination prefectures when the economy is booming. Therefore, the coefficients might reflect fathers' differential capacity to facilitate their children's access to employment rather than the heterogeneous impacts of housing prices.

Table 3 shows the estimation results using the instrumental variable approach. Row "K-P F-stat" shows the first-stage *F-stats*, which are much larger than 10, the rule of thumb, suggesting strong first stages. The estimation results of the first stages are shown in Appendix Table B.2. The impacts of the HPR policy on housing costs are negative, which is consistent with our expectations. The IV results are close to the OLS results. The negative coefficient of $HPgap$ suggests that higher housing price gaps decrease migration; the positive coefficient of the interaction term indicates that fathers' education can mitigate the deterring effect of housing costs. The IV estimations are slightly larger, but the differences between the OLS and IV results are not statistically significant.

The negative coefficient of $HPgap$ suggests that higher housing price gaps decrease income, though the coefficients are not statistically significant. The positive coefficient of the interaction term indicates that fathers' education can mitigate the adverse effect of housing costs. Again, the IV estimations are slightly larger, but the differences between the OLS and IV results are not statistically significant.

To interpret the magnitude of the coefficients, we first try to understand the magnitude of $HPgap$. $HPgap$ is a weighted average of the difference between log housing prices at the destinations and the log housing price at the origin in yuan per square meter. The mean value of $HPgap$ is .5, meaning that housing prices at the destinations are

Table 3: IV Results

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
FatherEdu× HPgap	0.013***	0.008***	0.031***	0.034***
	(0.002)	(0.003)	(0.011)	(0.012)
HPgap	-1.053***	-0.863**	-2.587	-2.701
	(0.377)	(0.366)	(2.054)	(2.047)
Edu× HPgap		0.015***		-0.009
		(0.003)		(0.015)
FatherEdu× birthyearFE	Y	Y	Y	Y
K-P F-stat	28.676	19.603	27.817	18.972
LM test	52.834	54.031	52.346	53.425
Obs.	14,050	14,050	11,308	11,308
Mean(Dep. Var.)	0.140	0.140	9.736	9.736
Mean(Dep. Var.) if FatherEdu = 0	0.083	0.083	9.123	9.123

NOTE: This table shows the main results 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. In Columns (1) and (3), the endogenous variables are $HPgap$ and $FatherEdu \times HPgap$; the instrumental variables are HPR and $FatherEdu \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEdu \times HPgap$, and $Edu \times HPgap$; the instrumental variables are HPR, $FatherEdu \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. 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 is 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.

about 165% of housing prices at the origin. We then calculate the impact of a 0.1 change in $HPgap$. As $HPgap$ is demeaned, the magnitude of the coefficient reflects the intergenerational persistence at the average housing price gap. The magnitude of $FatherEdu \times HPgap$ in Column (4) suggests that .1 increase in $HPgap$ can increase intergenerational persistence by 15%.[19] For comparison, Feigenbaum (2015) finds that a downturn during the Great Depression that is one standard deviation worse increases intergenerational persistence by 39%.[20]

As we are using cross-sectional data, the income of children is specifically for the year 2017. To address concerns that income data from a single year might reflect

[19] Although not shown in the table, we run regressions without controlling for the fathers' education times children's birth year fixed effects so that the coefficient can be estimated for $FatherEdu$, which is about 0.022. The calculation for the change in intergenerational persistence caused by a .1 change in $HPgap$ is 0.034×.1/0.022 = 15%.

[20] Feigenbaum (2015) reports that intergenerational persistence is 0.28 in a city experiencing an average Great Depression downturn. A one standard deviation increase in the severity of the downturn results in an additional 0.108 in persistence, corresponding to a 39% increase, calculated as 0.108/0.28*100%.

transitory shocks or measurement errors[21], we follow Nybom and Stuhler (2017) and average income across 2015, 2017, and 2019 waves of the CHFS, provided that the income data are available in these waves. Approximately 66% of observations include income information from more than one wave. Appendix Table B.3 displays results in Columns (3) and (4), where children's income is calculated as this multi-year average. For comparison, Columns (1) and (2) of the same table reproduce Columns (3) and (4) from the main results in Table 3. The estimated coefficients using both single-year and averaged income measures are very similar.

7.3 Robustness

7.3.1 IV Validity

Although we have controlled for employment and wages at the origin and destination, there may still be concerns that some unobservable economic conditions can potentially confound the results. To address this concern, we include origin × birth year fixed effects. The housing price gap is also at the origin times birth year level, where the potential destinations and the corresponding weights are specific to each origin. Therefore, the origin × birth year fixed effects would capture the time-variant characteristics in the origin and the origin-specific weighted destination shocks. The results are shown in Table 4 Columns (1), (2), (5), and (6). $HPgap$ is at the origin × birth year level, so they are absorbed by the fixed effects. The estimated coefficients of interest remain close to the main results.

Recent econometric literature (Feigenberg et al., 2023) has pointed out that, for interacted models, omitted variable bias may exist even when the level terms of confounders are controlled for. For this paper, one concern could be that the housing price gap is capturing the economic development at the destination. The impact of the economic development at the destination can be heterogeneous across children with different parental backgrounds. For example, children with more affluent fathers may respond more to economic growth at potential destinations due to information advantage. This concern can be addressed by controlling for the interaction term between fathers' education and economic conditions at the destination. In Table 4 Columns (3), (4), (7), and (8), we control for the interaction term between fathers' education and the employment, wage, fiscal expenditure, export, FDI at the destination prefectures, and the level terms of these measures of economic conditions. The results are close to the main results in Table 3

[21]This measurement error is not a big concern for this paper as income is the dependent variable, not the independent variable. The independent variable, fathers' education, is not affected by transitory shocks or measurement errors. Therefore, the

Table 4: Robustness: Additional Controls

	(1) Migration	(2) Migration	(3) Migration	(4) Migration	(5) lnIncome	(6) lnIncome	(7) lnIncome	(8) lnIncome
FatherEdu× HPgap	0.010***	0.008***	0.013***	0.008***	0.036***	0.039***	0.033***	0.037***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.013)	(0.013)	(0.011)	(0.012)
Edu× HPgap		0.006		0.015***		-0.011		-0.011
		(0.004)		(0.003)		(0.018)		(0.015)
HPgap			-0.942***	-0.794**			-1.489	-1.601
			(0.355)	(0.345)			(1.952)	(1.944)
OriginFE× BirthyearFE	Y	Y	-	-	Y	Y	-	-
Interaction terms	-	-	Y	Y	-	-	Y	Y
Obs.	14,050	14,050	14,050	14,050	11,308	11,308	11,308	11,308
Mean(Dep. Var.)	0.125	0.125	0.140	0.140	9.715	9.715	9.736	9.736
Mean(Dep. Var.) if FatherEdu = 0	0.077	0.077	0.083	0.083	9.108	9.108	9.123	9.123

NOTE: This table shows the results of adding more controls 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. In Columns (1) and (3), the endogenous variables are $HPgap$ and $FatherEdu \times HPgap$; the instrumental variables are HPR and $FatherEdu \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEdu \times HPgap$, and $Edu \times HPgap$; the instrumental variables are HPR, $FatherEdu \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level.

One potential concern is that affluent destinations are more likely to adopt the HPR policy. Additionally, the heterogeneous impacts of parental background may differ for origin prefectures that have more migrants moving to affluent areas. As a result, the estimated coefficient of interest might capture the heterogeneous impact of parental background due to specific origin characteristics that are correlated with higher migration rates to affluent destinations.

Although we have controlled for several origin characteristics, there remain unobservables that may be correlated with higher migration rates to affluent destinations and would potentially confound the results. To address this concern, we construct the indicators of whether a prefecture is one of the 46 prefectures that ever adopted the HPR policy. The indicators are then weighted in the same way as we do for the $HPgap$ and interact with fathers' education.

The results controlling for this interaction term are demonstrated in Table 5. Odd columns are copied directly from Table 3, and even columns are corresponding specifications controlling for the interaction term between the average weighted ever-treated indicators and fathers' education. The level term of the average weighted ever-treated indicators is at the origin level and thus absorbed by the origin fixed effects.

Table 5: Robustness: IV Validity

	(1) Migration	(2) Migration	(3) Migration	(4) Migration	(5) lnIncome	(6) lnIncome	(7) lnIncome	(8) lnIncome
FatherEdu× HPgap	0.013***	0.013***	0.008***	0.008***	0.031***	0.025**	0.034***	0.028**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.011)	(0.012)	(0.012)	(0.012)
HPgap	-1.053***	-1.075***	-0.863**	-0.886**	-2.587	-2.375	-2.701	-2.494
	(0.377)	(0.378)	(0.366)	(0.366)	(2.054)	(2.033)	(2.047)	(2.025)
Edu× HPgap			0.015***	0.015***			-0.009	-0.009
			(0.003)	(0.003)			(0.015)	(0.015)
FatherEdu× Evertreated		-0.005		-0.006		0.045		0.045
		(0.008)		(0.008)		(0.037)		(0.037)
K-P F-stat	28.676	29.279	19.603	20.003	27.817	28.490	18.972	19.420
Obs.	14,050	14,050	14,050	14,050	11,308	11,308	11,308	11,308
Mean(Dep. Var.)	0.140	0.140	0.140	0.140	9.736	9.736	9.736	9.736
Mean(Dep. Var.) if FatherEdu = 0	0.083	0.083	0.083	0.083	9.123	9.123	9.123	9.123

NOTE: This table shows the robustness results controlling for the average weighted ever-treated indicators and its interaction term with fathers' education. Fathers' years of education interacted with birth year fixed effects are controlled for in all columns. Odd columns are copied directly from Table 3, and even columns are corresponding specifications controlling for the average weighted ever-treated indicators and its interaction term with fathers' education. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

7.3.2 Drop Special Origin Prefectures

Characteristics of the origin prefecture may affect people's migration response to housing price changes. For example, 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 income effect. To check whether this possibility matters for our result, we run the main regressions excluding the prefectures that are adjacent to Beijing, Shanghai, Shenzhen, or Guangzhou.[22]

Additionally, the migration behaviors of people whose origin prefecture is economically more developed can differ from those of others. To check how much the characteristics of the origin matter for the results, we exclude from the sample the top ten prefectures in terms of GDP per capita in 2000 and run the main regressions.[23] Moreover, people in popular destinations may have different migration behavior from others. To examine the sensitivity of the main results to this aspect, the top 10 destinations, which unsurprisingly largely overlap with prefectures of high GDP per capita, are excluded from the sample.[24]

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

[23]The top ten prefectures in terms of GDP per capita in 2000 were Shenzhen, Shanghai, Zhuhai, Wuxi, Suzhou, Guangzhou, Beijing, Xiamen, Dongying, and Hangzhou. Other than megacities like Beijing or Shanghai, the top 10 prefectures vary in different years. Using the top 10 prefectures in other years gives very similar results.

[24]The top 10 most popular destinations are Beijing, Shanghai, Shenzhen, Guangzhou, Dongguan, Hangzhou, Suzhou, Chengdu, Ningbo, and Foshan, according to census sample data in 2010.

Table 6: Robustness: Sample Construction

	(1) Migration	(2) Migration	(3) Migration	(4) Migration	(5) lnIncome	(6) lnIncome	(7) lnIncome	(8) lnIncome
FatherEdu× HPgap	0.013***	0.012***	0.014***	0.014***	0.031**	0.030**	0.033**	0.034**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.012)	(0.012)	(0.014)	(0.014)
Exclude	-	adjacent to tier-1	high GDP per capita	popular destinations	-	adjacent to tier-1	high GDP per capita	popular destinations
Obs.	14,050	13,400	13,610	13,128	11,308	10,797	10,965	10,545
Mean(Dep. Var.)	0.140	0.142	0.143	0.146	9.736	9.727	9.710	9.681
Mean(Dep. Var.) if FatherEdu = 0	0.083	0.085	0.085	0.086	9.123	9.115	9.093	9.075

NOTE: This table shows the robustness results for different samples 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. In Columns (1) and (3), the endogenous variables are $HPgap$ and $FatherEdu \times HPgap$; the instrumental variables are HPR and $FatherEdu \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEdu \times HPgap$, and $Edu \times HPgap$; the instrumental variables are HPR, $FatherEdu \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table 6 shows the results restricting the origin prefectures as described above. Columns (1) and (5) copy the main 2SLS results in Table 3. Columns (2) and (6) demonstrate the estimations excluding the prefectures adjacent to four tier-1 prefectures. The estimations of the coefficients of interest change little, suggesting that the geographical spillover from the tier-1 prefectures is not a major concern for the results. Columns (3) and (7) demonstrate the estimations excluding origins with high GDP per capita. Columns (4) and (8) demonstrate the estimations excluding origins that are popular migration destinations. These estimations are also close to the main results.

7.3.3 Alternative Housing Price Dataset

The main dataset we use for housing prices is from the CEIC database. To check if the patterns we observe are specific to the CEIC dataset, we use a different dataset of housing prices from Anjuke. Anjuke is a prominent online real estate platform in China, recognized for its extensive listings and comprehensive services related to property buying, selling, and renting. As one of the largest and most influential entities in the Chinese real estate market, Anjuke provides a vast array of real estate information.

The results are shown in Table 7. The estimations are close to the main results using the CEIC data, suggesting that we are not capturing dataset-specific patterns.

Table 7: Robustness: Alternative Housing Price Dataset

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
FatherEdu× HPgap	0.013***	0.007**	0.036***	0.046***
	(0.003)	(0.003)	(0.012)	(0.013)
HPgap	-2.236***	-1.855***	-2.880	-3.539
	(0.657)	(0.614)	(2.735)	(2.741)
Edu× HPgap		0.017***		-0.030*
		(0.004)		(0.016)
FatherEdu× birthyearFE	-	Y	-	Y
K-P F-stat	19.184	13.397	19.163	13.150
Obs.	14,026	14,026	11,291	11,291
Mean(Dep. Var.)	0.138	0.138	9.737	9.737
Mean(Dep. Var.) if FatherEdu = 0	0.081	0.081	9.125	9.125

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. In Columns (1) and (3), the endogenous variables are $HPgap$ and $FatherEdu \times HPgap$; the instrumental variables are HPR and $FatherEdu \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEdu \times HPgap$, and $Edu \times HPgap$; the instrumental variables are HPR, $FatherEdu \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

7.3.4 Alternative Measure of Fathers' Education

Because our dataset consists of children from different prefectures who were born in different years, the same year of fathers' education may reflect different parental backgrounds. For example, fathers living in remote areas with high school education can be very different from fathers with high school education in Beijing in terms of socioeconomic status. To address this concern, I replace the original measure with the residualized fathers' years of education, which is obtained by regressing fathers' years of education on father's *Hukou* status and origin prefecture times children's birth year fixed effects.

The results are shown in Table 8. The estimations are similar to the main results in Table 3, suggesting that the results are not sensitive to the measurement of fathers' education.

Table 8: Robustness: Residualized Fathers' Education

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
FatherEdu_r× HPgap	0.006**	0.006**	0.025**	0.028**
	(0.003)	(0.003)	(0.013)	(0.013)
HPgap	-0.841**	-0.842**	-0.366	-0.349
	(0.401)	(0.401)	(2.301)	(2.303)
Edu× HPgap		0.004		-0.012
		(0.004)		(0.018)
FatherEdu_r× birthyearFE	Y	Y	Y	Y
K-P F-stat	26.532	17.687	25.103	16.730
LM test	47.957	47.958	47.060	47.051
Obs.	12,858	12,858	10,324	10,324
Mean(Dep. Var.)	0.125	0.125	9.715	9.715

NOTE: This table shows the results using residualized fathers' years of education to measure parental background, using the HPR policy treatment as the instrumental variable. 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. 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 $FatherEdu\_r \times HPgap$; the instrumental variables are HPR and $FatherEdu\_r \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEdu\_r \times HPgap$, and $Edu \times HPgap$; the instrumental variables are HPR, $FatherEdu\_r \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

7.3.5 Impute Father's Income and Household Wealth

While education is an important proxy of parental background, it is far from being the only one. Fathers with the same educational attainment can still have very different incomes and levels of household wealth. To address this concern, we replace fathers' education with imputed fathers' income and household wealth in the following robustness tests.

To address the lifecycle bias and transitory shocks in income, researchers commonly employ proxies such as parental education or occupation to proxy parental lifetime income (Solon, 1992; Gong et al., 2012). In line with this methodology, I utilize parental education, *Hukou* status, Communist party membership, job position, and coastal region indicators to impute parental lifetime income. These variables are relatively stable throughout the life cycle and strongly correlate with lifetime earnings, thereby reducing lifecycle bias. Moreover, the use of estimated income (rather than actual income) helps to decrease the attenuation bias caused by temporary income fluctuations (Fan et al., 2021).

Specifically, the imputation takes the following steps. First, we use CHFS 2015 data, restrict the sample to males, and estimated the following equation:

$$ln(Income)_i = \alpha_0 + \alpha_1 Age_i + \alpha_2 Age\_squared_i + X_i' \alpha_X + \epsilon_i \tag{10}$$

X is a comprehensive 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.[25] Then, based on Equation (10), we compute income for our main sample using information on their fathers' characteristics and the estimated coefficients.[26]

Replacing fathers' education with the imputed fathers' income, we re-estimate the main results. Table 9 shows the results. The F-stats of the first stages remain much larger than 10, suggesting that we have strong first stages. The coefficients of $HPgap$ are negative across all columns, indicating that housing costs are a barrier to migration, and an increase in housing costs decreases income. The coefficients of the interaction term, $FatherIncome \times HPgap$, are positive and significant in all columns, suggesting that fathers' income can mitigate the negative impacts of housing costs.

To evaluate the goodness of the imputation, we compare our estimation of the intergenerational income elasticity (IGE) to the existing literature. We show results in Appendix Table B.4 where we dropped the $FatherIncome \times birthyear$ fixed effects to directly observe the estimated coefficient of fathers' log income, i.e., the estimation of IGE. Our IGE estimation 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

[25]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, we separate fathers into 10-year age groups based on their birth years. When the father's birth year is not available, we use the average birth year of fathers given the child's birth year, which we generated using the 2013 China Health and Retirement Longitudinal Study (CHARLS) data, a nationally representative dataset focusing on the old population.

[26]We do not use the coefficients of age, age squared, or birth cohort because we are 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 method entails calculating the imputed income under the assumption that all individuals are at the mean age, thereby adding a constant across all observations. The results will be the same, as the imputed income of fathers is centered (demeaned) prior to estimating the regression models. We also checked the results with age-related variables, and they remain robust.

Table 9: Robustness: Impute Fathers' Income

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
lnFatherIncome× HPgap	0.211***	0.122***	0.514***	0.548***
	(0.041)	(0.045)	(0.178)	(0.191)
HPgap	-1.025***	-0.814**	-2.682	-2.764
	(0.387)	(0.373)	(2.064)	(2.057)
Edu× HPgap		0.016***		-0.006
		(0.003)		(0.015)
lnFatherIncome× birthyearFE	Y	Y	Y	Y
K-P F-stat	28.355	19.452	27.786	19.016
LM test	51.970	53.310	51.857	53.096
Obs.	13,925	13,925	11,211	11,211
Mean(Dep. Var.)	0.139	0.139	9.736	9.736

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 $Edu \times HPgap$; the instrumental variables are HPR, $FatherIncome \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. 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 is 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.

birth cohort.[27][28]

A similar procedure is applied to impute household assets. Appendix Table B.5 shows the results where fathers' education is replaced with imputed household assets. The results remain robust.

[27]The measurement and sample period in our 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 our sample includes individuals born between 1962-2000.

[28]I acknowledge the presence of selection bias regarding whose incomes are documented in our dataset. Specifically, we lack income data for temporary migrants and deceased individuals. Fan et al. (2021), which aims to precisely estimate the magnitude of intergenerational income elasticity (IGE), suggests that adjustments for this selection bias do not yield statistically significant differences in the IGE estimations.

7.4 Heterogeneous Analysis

7.4.1 Fathers' Education Levels

This section investigates heterogeneous impacts across different demographic groups. First, we check which level of father's education matters most for overcoming housing price barriers.

Table 10: Heterogeneity: Fathers' Education Levels

	OLS		IV	
	(1) Migration	(2) lnIncome	(3) Migration	(4) lnIncome
$1\{\text{FatherEduType} \geq 2\} \times$ HPgap	0.032***	0.138*	0.058***	0.242**
	(0.012)	(0.073)	(0.021)	(0.123)
$1\{\text{FatherEduType} \geq 3\} \times$ HPgap	0.029*	0.077	0.023	0.128
	(0.017)	(0.074)	(0.023)	(0.096)
$1\{\text{FatherEduType} \geq 4\} \times$ HPgap	-0.008	0.008	0.003	0.035
	(0.022)	(0.082)	(0.028)	(0.106)
HPgap	-0.360***	-0.816**	-0.758**	-2.391
	(0.078)	(0.386)	(0.370)	(2.077)
K-P F-stat			11.520	11.051
Obs.	14,050	11,308	14,050	11,308
Mean(Dep. Var.)	0.140	9.736	0.140	9.736
SD(Dep. Var.)	0.346	1.683	0.346	1.683

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

Table 10 shows the results interacting different levels of fathers' education attainment with the housing price gaps. Father's education attainment, $FatherEduType$, is grouped into four levels: no schooling (32.36% of the sample), primary school (32.91%) 3) junior high school (19.21%), and senior high or above (15.52%). In the regression, "no schooling" is taken as the reference group. To capture the additional impact for each education level, we use the indicators for having a certain education level and above. For example, the coefficient of the interaction term $1\{FatherEduType \geq 3\} \times HPgap$ reflects the additional impact of having a junior high school education, compared to only a primary school education, in overcoming housing barriers. A significant and positive coefficient means that additional education helps. In contrast, an insignificant coefficient implies that beyond previous educational attainment, the additional education level does not help overcome the housing price barrier.

Columns (1) and (3) in Table 10 show the impacts on migration. The results suggest that, compared to illiterate fathers, it is mainly whether the father has a primary school education that helps with overcoming housing costs. A junior high school education or above does not provide additional benefits in dealing with housing expenses. Columns (2) and (4) show the impacts on income. Consistent with the results on migration, whether the father has a primary school education matters most for overcoming the negative effects of housing costs on income.

7.4.2 Male vs Female

For gender heterogeneity, we observe the impacts of father's education on migration decisions for both males and females in Columns (1) and (4) in Table 11. However, the impact on income is only significant for males, according to Columns (2) and (4). As we only include observations with available income information, the impact may be more pronounced on spouses' income for females. In Columns (3) and (6), the outcome is the log income of the spouse. The magnitude is not small but noisily estimated.

Table 11: Heterogeneity: Male vs Female

	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
	Migration	lnIncome	lnIncome_s	Migration	lnIncome	lnIncome_s
FatherEdu× HPgap	0.006*	0.052***	0.028	0.009***	0.004	0.019
	(0.004)	(0.015)	(0.018)	(0.003)	(0.015)	(0.013)
K-P F-stat	13.452	15.052	14.917	20.165	17.035	20.240
Obs.	6,461	5,700	4,408	7,589	5,608	5,847
Mean(Dep. Var.)	0.140	10.045	9.576	0.139	9.422	10.072
Mean(Dep. Var.) if FatherEdu = 0	0.073	9.527	9.077	0.091	8.750	9.612

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

7.4.3 Rural vs Urban

We also check heterogeneity across children with parents of different *Hukou* types. We did not use children's *Hukou* type directly as children's *Hukou* type is correlated to their migration behavior. Columns (1) and (2) in Table 12 show the results for children with any rural *Hukou* parent, and Columns (3) and (4) show the results for children with urban *Hukou* parents. The results suggest that the impact of parental backgrounds in overcoming housing cost barriers mainly matters for children with rural *Hukou* parents. The differences in the coefficients of interest are statistically significant. The results indicate that it is the rural-urban migration that is affected most by the housing cost barriers.

Table 12: Heterogeneity: Rural vs Urban

	Rural		Urban	
	(1)	(2)	(3)	(4)
	Migration	lnIncome	Migration	lnIncome
FatherEdu× HPgap	0.013***	0.048***	0.003	-0.015
	(0.003)	(0.017)	(0.004)	(0.015)
K-P F-stat	14.174	13.701	17.496	16.211
Obs.	9,442	7,767	4,608	3,541
Mean(Dep. Var.)	0.146	9.438	0.127	10.391
Mean(Dep. Var.) if FatherEdu = 0	0.085	9.005	0.075	9.805

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

7.5 Impacts on Structural Transformation

In the heterogeneous analyses, we find the impacts more pronounced for people with rural *Hukou* parents. As migration is often associated with non-agricultural employment, housing prices may have varied effects on agricultural employment across different parental backgrounds, which indicates an interaction between housing costs, structural transformation, and intergenerational mobility. To investigate this interaction, we examine the impacts of housing costs on agricultural employment.

We replace the outcome of the main analyses with the indicator for whether the individual works in the agricultural sector. The results are shown in Table 13. Columns (1) and (2) show the results for the full sample, and Columns (3) and (4) show the results for children with a parent of rural *Hukou*. The results imply that housing costs increase the probability of working in the agricultural sector, and higher fathers' education decreases this impact. The heterogeneous effects of housing costs on agricultural work across

Table 13: Impacts on Structural Transformation

	(1) Agri_sector	(2) Agri_sector	(3) Agri_sector	(4) Agri_sector
FatherEdu× HPgap	-0.012***	-0.013***	-0.011***	-0.012***
	(0.002)	(0.003)	(0.004)	(0.004)
HPgap	1.456***	1.467***	1.595**	1.618**
	(0.518)	(0.521)	(0.777)	(0.785)
Edu× HPgap		0.001		0.002
		(0.004)		(0.005)
FatherEdu× birthyearFE	Y	Y	Y	Y
K-P F-stat	29.202	19.705	21.311	14.280
Obs.	11,167	11,167	7,582	7,582
Mean(Dep. Var.)	0.226	0.226	0.315	0.315
Mean(Dep. Var.) if FatherEdu = 0	0.370	0.370	0.419	0.419

NOTE: This table shows the results using agricultural employment as the outcome. 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 $FatherEdu \times HPgap$; the instrumental variables are HPR and $FatherEdu \times HPR$. In Columns (2) and (4), the endogenous variables are $HPgap$, $FatherEdu \times HPgap$, and $Edu \times HPgap$; the instrumental variables are HPR, $FatherEdu \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

parental backgrounds suggest that housing costs can affect structural transformation for the economy, interacting with inequality and intergenerational persistence.

8 Spatial Equilibrium Model, Counterfactual, and Policy Evaluations

In the regression analysis, we aggregate the impacts of housing prices at potential destinations to a weighted average, with the outcome being a binary decision on whether to migrate or not. In reality, migration is not a binary choice but a choice set with different potential destinations, where the housing price at one destination may affect the migration flows to the other destinations. By developing a spatial equilibrium model, we are able to model the migration decision as a discrete choice and analyze cross-price elasticity, capturing this complexity.

While the current analyses focus on the household side and do not discuss the impacts of migration on the destinations, migration certainly has general equilibrium effects. For example, when housing prices decrease, migration increases; housing prices at the destination will then be endogenously driven up by the inflow of migrants, preventing further migration.

Through counterfactual analyses, the spatial equilibrium model can be used to simulate the outcomes of various policy interventions, such as housing subsidies, changes in land use regulations, or increases in housing supplies in high-demand areas for migrants. Additionally, it enables us to evaluate how the policy impacts would differ depending on which destination prefectures and which subset of populations they target. By comparing the simulated outcomes with the current equilibrium, the model can provide valuable insights into the most effective strategies for promoting greater economic mobility through enhanced housing affordability. Moreover, we would be able to assess the extent to which intergenerational persistence, geographical inequality, and barriers to structural transformation may be reduced with various housing policies.

Another benefit of constructing a spatial equilibrium model, which is particularly relevant in the Chinese context, is that it allows for a quantitative comparison of the differential impacts of *Hukou* restrictions versus housing barriers.

In this section, we develop a spatial equilibrium model to generate counterfactuals. The model follows a standard framework, with one key tweak: it incorporates a migration cost related to housing prices, with impacts that vary based on parental backgrounds.

8.1 Setup

8.1.1 Labor Supply / Worker's Choice of Destination and Housing Demand

Each worker, indexed by i, chooses to live in the prefecture that offers him the most desirable bundle of wages, housing prices, and moving costs. A worker living in prefecture p inelastically supplies one unit of labor and earns a wage of w_p. Living in a prefecture, the worker consumes housing M, which has a local price of rR_p, where r is the discount factor that transfers housing purchase prices to rental prices, and a national good O, which has a national price that we normalize to one. The worker has Cobb-Douglas preferences for housing and national good, which they maximize subject to their budget constraint:

$$\max_{M,O} \quad \ln\left(M^{\zeta}\right) + \ln\left(O^{1-\zeta}\right)$$

$$\text{s.t.} \quad O + rR_pM \leq W_p.$$

Workers' relative taste for national versus local goods is governed by ζ, where $0 \leq \zeta \leq 1$. ζ is assumed to be constant across workers. The worker's optimized utility function from consumption can be expressed as an indirect utility function for living in prefecture p. If the worker were to live in prefecture p, his indirect utility from consumption, V_p, would be:

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

where $w_p = ln(W_p)$ and $h_p = ln(R_p)$. The worker's optimized utility function also leads to his housing demand (D_p):

$$D_p = \frac{\zeta W_p}{rR_p} \tag{11}$$

Additionally, the worker pays moving costs C_{iop}:

$$C_{iop} = \beta^h_{fedu(i)} h_p + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p\} + \beta^{Dis} lnDistance_{op}$$

where $\beta^h_{fedu(i)} h_p$ captures the transient residential cost that occurs when the worker searches for a job once they arrive at the destination. The impacts of h_p differentiate across parental backgrounds as workers from richer families are more capable of paying the upfront residential cost. $\beta^{ruralHK}$ measures the additional utility for workers from rural areas if they migrate. Rural workers may derive greater utility from migrating than their urban counterparts, as they often access significantly improved job opportunities and amenities compared to what is available in their rural hometowns. $\beta^{Dis} lnDistance_{op}$ captures the transportation cost or the utility cost of being far away from home.

Additionally, there is destination-specific attractiveness or disutility that we capture using destination fixed effects, μ_p. Moreover, each worker has an idiosyncratic taste for each prefecture, which is measured by ϵ_{iop}. ϵ_{iop} is drawn from a Type I Extreme Value distribution, with dispersion parameter σ. Combined, the utility a worker obtains by choosing to live in prefecture p is:[29] [30]

$$\begin{aligned} U_{iop} = (w_p - \zeta h_p) + \beta^h_{fedu(i)} h_p + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p\} \\ + \beta^{Dis} lnDistance_{op} + \mu_p + \epsilon_{iop} \end{aligned} \tag{12}$$

[29]We use the conditional logit model. The IIA may not hold in some cases. For example, assume prefectures A and B are the same except for the distance to Beijing. Beijing's housing price changes should affect the migration to the cheaper prefecture A, which is close to it, but may not affect the migration to a remote, more affordable prefecture B. The relative migration between A and B will be different, given the changes in Beijing, but the current model does not allow it. We may address this concern later, using nested logit or mixed logit with random coefficients.

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

We rewrite this equation as

$$U_{iop} = \delta_p^{fedu} + \beta^{Dis} lnDistance_{op} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p\} + \epsilon_{iop}$$

where $\delta_p^{fedu} = (w_p - \zeta h_p) + \beta^h_{fedu(i)} h_p + \mu_p$.[31]

$lnDistance_{op}$ and $ruralHukou_i$ are taken as exogenous while h_p and w_p are endogenous. This setup is the conditional logit model, first formulated in this utility maximization context by McFadden (1973). Aggregate population differences of workers across prefectures represent differences in these workers' mean utility values for these prefectures. The total expected population of prefecture p is the probability each worker lives in the prefecture, summed over all workers:

$$L_p = \sum_{i \in L} \frac{exp(\delta_p + \beta^{Dis} lnDistance_{op} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p\})}{\sum_{n \in N} exp(\delta_n + \beta^{Dis} lnDistance_{on} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq n\})}$$

where L is the set of workers in the nation.

The key difference from the standard framework in this setup is how h_p is treated. Specifically, h_p enters as both part of the consumption utility and the moving cost. Only the consumption expenditure on housing affects the housing demand. The impacts of housing prices on moving costs do not affect housing demand. The intuition is that the effect of housing prices on moving costs is 'transient' and, particularly, fathers' education only affects this transient moving cost but not the long-term consumption of housing. This fits what is in reality most properly.

For robustness, we can experiment with some alternative setups. For example, instead of the transient migration cost due to housing prices at the destination, a worker with different parental backgrounds has different preferences for housing. Namely, instead of ζ, we should use ζ_{fedu}. As a result, workers from different parental backgrounds spent a different share of income on housing. We then drop the h_p in moving costs C_{iop} and get indirect utility function:

$$\begin{aligned} U_{iop} = (w_p - \zeta_{fedu(i)} h_p) + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p\} \\ + \beta^{Dis} lnDistance_{op} + \mu_p + \epsilon_{iop} \end{aligned} \tag{13}$$

[31] If not too computationally burdensome, we can replace μ_p with $\mu_p + \mu'_p \times \mathbf{1}\{p \neq o\}$ in case native and migrants have different region-specific utilities. Also, we may want to add one more term to δ_{op}, which is origin times destination fixed effects. The fixed effects can capture some preference that is specific to the origin-destination pair, e.g., social network. Alternatively, the impact of social network can be directly added to δ_{op} with imposed structure $\beta^{net} network_{op} + \beta^{Hometown} * \mathbf{1}\{p = o\}$ where $network_{op} = 0$ if $p = o$.

Alternatively, instead of $(w_p - \zeta h_p)$, the utility from consumption may be heterogeneous across parental background, $(w_p - \zeta h_p) \times \phi_{fedu}$. This suggests heterogeneous utility from consumption, which is adopted in Diamond (2016).

8.1.2 Labor Demand

We assume that within each prefecture, firms produce a homogenous tradeable good and share identical production technology with constant returns to scale production functions. Therefore, firm-level labor demand translates directly to prefecture-level aggregate labor demand. For simplicity of notation, we assume that each prefecture has a representative firm that maximizes profits. Let the price of output be normalized to one. In each prefecture p, the representative firm takes productivity A_p as given and produces goods according to the production function:

$$Y_p = A_p L_p$$

where L stands for labor in the prefecture, and K stands for capital. Note that, to keep the notation simple, we assume that natives and migrants are perfect substitutes at the prefecture level.[32]

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:

$$w_p = lnA_p$$

8.1.3 Housing Supply

Following Diamond (2016), housing prices, h_p, are determined by the equilibrium in the housing market. The production of housing relies on inputs such as construction materials and land. Developers act as price-takers and sell homogenous houses at the marginal cost of production.

$$h_p = MC(\kappa_p, \eta_p)$$

The function $MC(\kappa_p, \eta_p)$ represents the relationship between local construction costs, κ_p, and local land costs, η_p, and the marginal cost of constructing a home. The cost of land η_p is a function of the aggregate demand for housing. I parameterize the

[32] We can introduce imperfect substitutability between natives and migrants by assuming that $L_p = ((\sum_{p \neq o} L_{op})^{\rho} + L_{pp}^{\rho})^{\frac{1}{\rho}}$, where L_{op} indicates migrants moved from origin o to location p and L_{pp} indicates natives in location p.

log housing supply equation as follows:

$$h_p = ln(\kappa_p) + \gamma ln(D_p) \tag{14}$$

The elasticity of housing prices γ with respect to housing demand will be calibrated based on the existing literature.

8.2 Equilibrium

Equilibrium in this model is defined by a menu of wages and housing prices $(w_p,\ h_p)$ with populations (L_p), such that:

- The labor demand equals labor supply:

$$L_p = \sum_{i \in L} \frac{exp(\delta_p^{fedu} + \beta^{Dis} lnDistance_{op} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p\})}{\sum_{n \in N} exp(\delta_n^{fedu} + \beta^{Dis} lnDistance_{on} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq n\})}$$

where $\delta_p^{fedu} = \beta^w (w_p - \zeta h_p) + \beta^h_{fedu(i)} h_p + \mu_p + \xi_p$.

$$w_p = lnA_p$$

- Housing demand equals housing supply:

$$h_p = ln(\kappa_p) + \gamma ln(D_p)$$

$$D_p = L_p \frac{\zeta exp(w_p)}{rexp(h_p)}$$

The model does not allow me to solve for equilibrium wages and housing prices analytically, but this setup is useful in estimation.

8.3 Estimation and Simulations

8.3.1 Labor Supply

Before we analyze the general equilibrium, we first estimate only the labor supply side. The estimation results can be used to do back-of-envelope calculations and evaluate policy changes.

The setup is the conditional logit model, as formulated by McFadden (1973). The utility that a worker i from origin o born in year t obtains by choosing to live in prefecture

p is given by:

$$U_{iopt} = \delta_{pt}^{fedu} + \beta^{Dis} lnDistance_{op} + \beta^{ruralHK} ruralHuKou_i \times 1\{o \neq p\} + \epsilon_{iopt} \quad (15)$$

where δ_{pt}^{fedu} is defined as:

$$\delta_{pt}^{fedu} = \beta_{fedu} h_{pt} + \beta^{w} w_{pt} + \mu_p + \pi_{fedu,t} + \xi_{pt}$$

and ϵ_{iopt} is drawn from a standard Type I Extreme Value distribution. Here, $fedu$ is the worker's father's education level. The magnitudes of the coefficient on wages and housing prices represent the elasticity of workers' demand for a small prefecture with respect to its local wages and housing prices, respectively. [33]

I estimate the model in two steps. First, I estimate Equation (15) using MLE. Secondly, I will use the estimated $\hat{\delta}_{pt}^{fedu}$ and apply a 2SLS estimation using HPR policy as the instrument for housing prices h_{pt} to estimate β_{fedu}^{h}.

The log-likelihood of the first step is:

$$l = \sum_i \log \left(\frac{\exp\left(\delta_{p(i)t}^{fedu} + \beta^{Dis} lnDistance_{op(i)} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq p(i)\}\right)}{\sum_{n \in N} \exp\left(\delta_{nt}^{fedu} + \beta^{Dis} lnDistance_{on} + \beta^{ruralHK} ruralHukou_i \times 1\{o \neq n\}\right)} \right)$$

where N is the set of all destination prefectures, and $p(i)$ is the destination chosen by worker i.

In practice, I do not have enough data to estimate $\delta_{p(i)t}^{fedu}$ for each birth year and each father's education level. Instead, I separate the sample into early and late birth cohorts based on birth years and fathers' education into high and low. In this way, I estimate 4 δ's for each prefecture. Then I estimate the following equation using 2SLS estimation:

$$\begin{aligned} \hat{\delta}_{pt}^{fedu} &= \beta_1 FatherEduHigh \times HousingPrices_{pt} + \beta_2 HousingPrices_{pt} \\ &+ M_{pt}'\Theta + \mu_p + \pi_{fedu,t} + \xi_{pt} \end{aligned} \quad (16)$$

where $FatherEduHigh$ is an indicator for fathers' education being high type. M_{pt} are macroeconomic conditions including employment rate, GDP per capita, log wage, and log export. μ_p are prefecture fixed effects, and $\pi_{fedu,t}$ are fathers' education times birth

[33] Due to the functional form assumption for the distribution of workers' idiosyncratic tastes for cities, the elasticity of demand of workers with father's education $fedu$ and birth cohort t for a prefecture p with respect to local housing prices, for example, is $(1 - s_{pt})\beta_{fedu}^{h}$. s_{pt} is the share of all workers of birth cohort t, living in prefecture p. For a small city, where the share of all workers of birth cohort t living in prefecture p is close to zero, the demand elasticity for rent is simply β_{fedu}^{h}.

cohort fixed effects. ξ_{pt} is the residual including local amenities. Housing prices can be correlated with local amenities and are, therefore, endogenous. I instrument housing prices using the HPR policy as in the main analyses.

The results of the second step are shown in Table 14. Columns (1) and (2) show the OLS estimates, both without and with macroeconomic controls. Columns (3) and (4) show the IV estimates, again with and without these controls. The coefficients on housing prices are negative and significant, indicating that higher housing prices at the prefecture reduce a worker's utility of residing there, making the prefecture less attractive as a migration destination. The coefficients for the interaction term between fathers' education and housing prices are positive and significant, suggesting that higher levels of fathers' education can partially mitigate the negative effects of housing prices. However, this mitigation is insufficient to fully counterbalance the adverse impacts of high housing prices.

The results of the first stage of the 2SLS estimation are shown in Appendix Table B.6. The coefficients exhibit the expected signs. Specifically, the coefficients associated with the HPR policy, when housing price is the independent variable, are negative. This indicates that the implementation of the HPR policy leads to a reduction in housing prices

Table 14: Conditional logit

	OLS		IV	
	(1)	(2)	(3)	(4)
FatherEduHigh× HousingPrices	0.381***	0.382***	0.730**	0.729**
	(0.110)	(0.110)	(0.327)	(0.327)
HousingPrices	-1.218***	-1.011***	-4.138***	-3.567**
	(0.355)	(0.376)	(1.276)	(1.447)
FatherEduHigh× birthcohort FE	Y	Y	Y	Y
Prefecture FE	Y	Y	Y	Y
Controls	-	Y	-	Y
K-P F-stat			12.370	7.772
LM test			18.561	9.569
Obs.	603	603	603	603
Mean(Dep. Var.)	0.168	0.168	0.168	0.168

NOTE: SEs are clustered at the prefecture level. Columns (1) and (2) show the OLS estimates, both without and with macroeconomic controls including 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 results are robust when the first step is made more complex by adding the worker's education and social networks between origin and destination and by adjusting the thresholds for father's education and birth cohorts.

(This section is still in progress.)

9 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.

This paper proposes that in China, a country characterized by geographical inequality (Xie and Zhou, 2014) and high intergenerational persistence, 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. For future research, comparative studies across countries with varying housing markets and migration policies could offer a broader perspective on the global implications of housing affordability for intergenerational mobility.

In conclusion, this research underscores the critical need for targeted policies to alleviate housing affordability issues and promote economic equality and mobility. As housing markets continue to evolve, ongoing research and policy innovation will be essential to mitigating the economic and social impacts of housing affordability.

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). 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.

Specifically, we assume that if a child stays at the origin, they get $C_{stay} = \mu_0 + \delta_0 F$; if the children migrate, they get $C_{migrate} = \mu_1 + \delta_1 F$. Migration requires a cost of $M(h, F)$, which depends on the housing price gap h and the father's socioeconomic status F. We 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 we call "disutility" and denote as τ. τ is independent of F and the realized C. We denote the CDF and PDF function of τ as $F_\tau(\tau)$ and $f_\tau(\tau)$

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

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

$$C_{migrate} - M(h, F) - \tau > C_{stay} \tag{17}$$

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

$$(\mu_1 - \mu_0) + (\delta_1 - \delta_0)F - M(h, F) > \tau \tag{18}$$

Denote the lefthand 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 our main regression analysis, we 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, we derive the cross derivative of the expected income with respect to h and F:

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

B Appendix: Figures and Tables

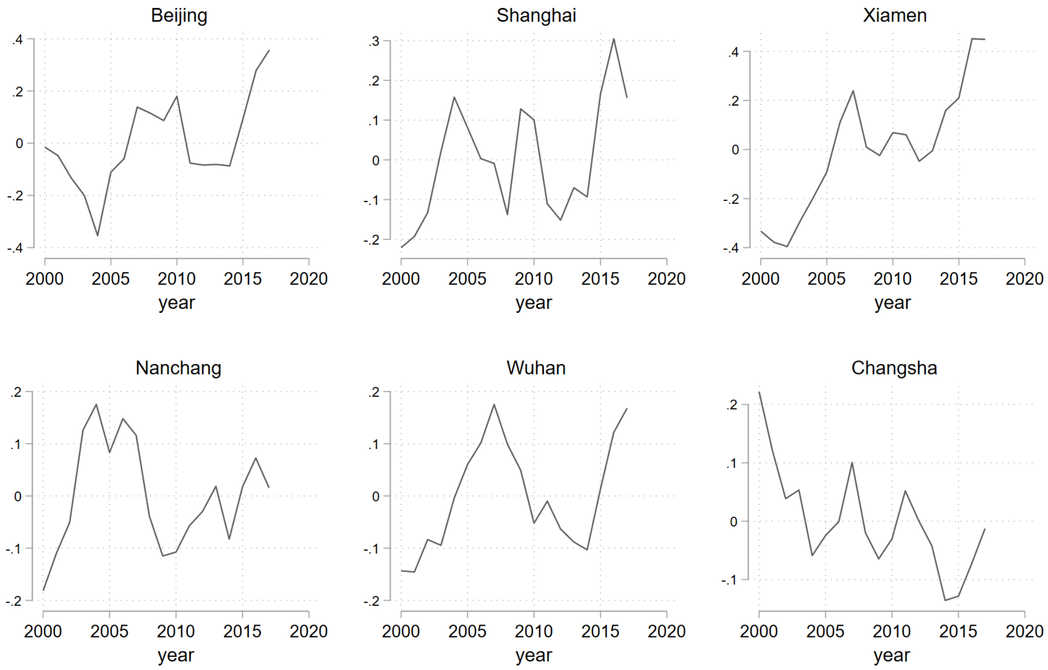


Figure B.1: Housing price residuals in six major prefectures in China

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 variations across prefectures.

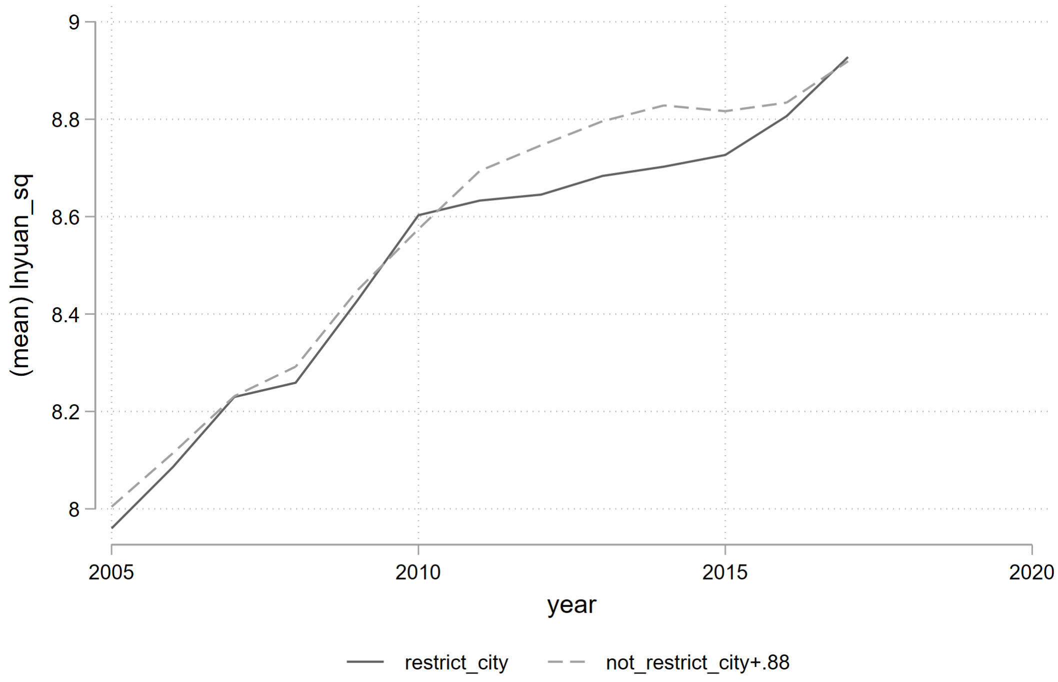


Figure B.2: Housing Prices for HPR and non-HPR prefectures

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 clearer, the housing prices of the never-treated prefectures are shifted upward by 0.88 log points.

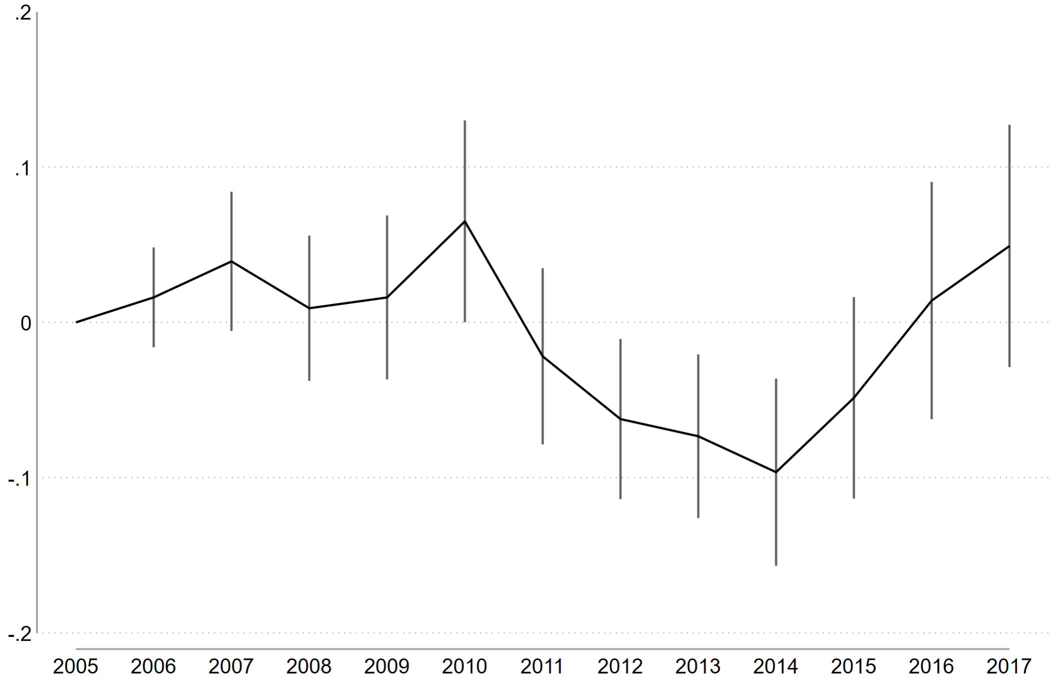
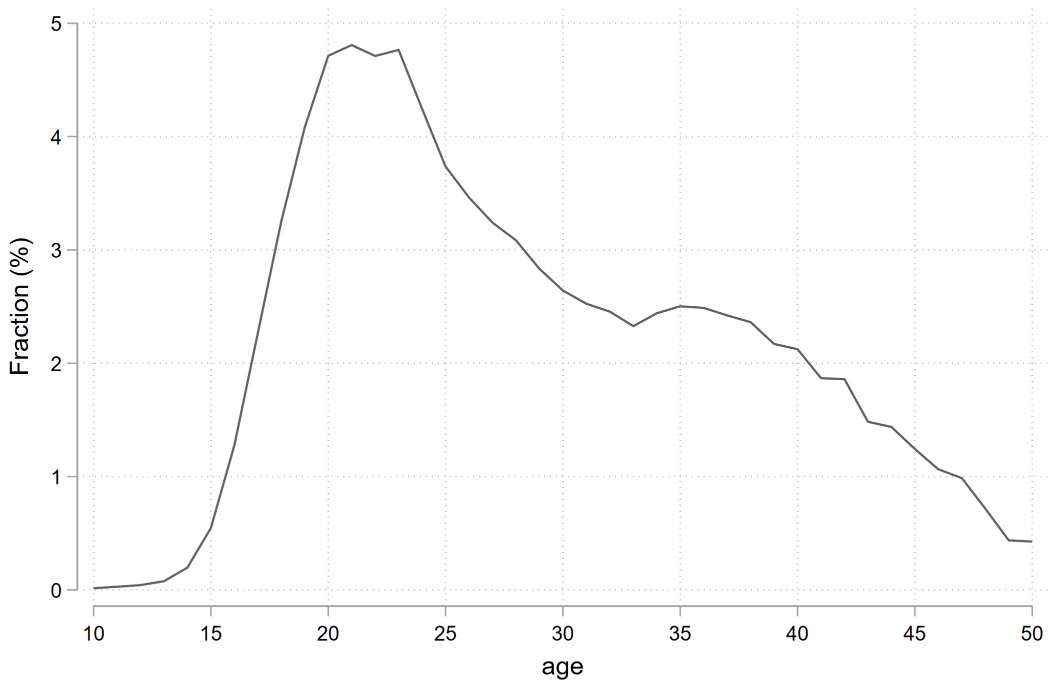


Figure B.3: HPR Event Study Plot

NOTE: The figure shows the event study plot for the policy. Specifically, 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 having adopted the policy. The figure shows the estimations 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.4: 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.

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 from Chen et al. (2017)

Table B.2: IV Results: First Stage

	(1) HPgap	(2) FatherEdu× HPgap	(3) HPgap	(4) FatherEdu× HPgap
HPR	-0.131***	-12.706***	-0.126***	-12.935***
	(0.018)	(0.595)	(0.018)	(0.602)
FatherEdu× HPR	0.000	2.512***	-0.001	2.559***
	(0.001)	(0.066)	(0.001)	(0.067)
FatherEdu× birthyearFE	-	Y	-	Y
Edu× birthyearFE	-	-	Y	Y
Obs.	14,050	14,050	14,050	14,050
Adj. R-sq	0.995	0.820	0.995	0.820

NOTE: The figure shows the first-stage estimations. Note that the IV Probit model uses MLE for estimation rather than 2SLS. The estimates in the table when the outcome is migration, however, is the standard first stage that we do in 2SLS estimations. Therefore, these results are just for illustrative purposes, showing that we have a strong first stage. When included, Edu×HPgap is also instrumented using Edu×HPR. Standard errors are clustered at prefecture times birth year level; ***p<0.01, **p<0.05, *p<0.1.

Table B.3: Robustness: Average Income Across Waves

	(1) lnIncome	(2) lnIncome	(3) lninc_avgwaves	(4) lninc_avgwaves
FatherEdu× HPgap	0.031***	0.034***	0.030***	0.034***
	(0.011)	(0.012)	(0.010)	(0.011)
HPgap	-2.587	-2.701	-3.008	-3.153*
	(2.054)	(2.047)	(1.902)	(1.902)
Edu× HPgap		-0.009		-0.011
		(0.015)		(0.013)
FatherEdu× birthyearFE	Y	Y	Y	Y
K-P F-stat	27.817	18.972	26.726	18.216
LM test	52.346	53.425	49.881	50.902
Obs.	11,308	11,308	12,045	12,045
Mean(Dep. Var.)	9.736	9.736	9.256	9.256
Mean(Dep. Var.) if FatherEdu = 0	9.123	9.123	8.651	8.651

NOTE: This table presents results using the log of average income across waves ($lninc\_avgwaves$), 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 (3) and (4). 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.4: Intergenerational Income Elasticity (IGE)

	(1)	(2)
	lnIncome	lnIncome
lnFatherIncome	0.382***	0.384***
	(0.062)	(0.062)
lnFatherIncome× HPgap	0.515***	0.550***
	(0.191)	(0.202)
HPgap	-2.351	-2.431
	(1.960)	(1.956)
Edu× HPgap		-0.006
		(0.015)
K-P F-stat	30.459	20.761
LM test	56.696	57.841
Obs.	11,211	11,211
Mean(Dep. Var.)	9.736	9.736

NOTE: This table shows the results using imputed fathers' income in place of fathers' education. In Columns (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 $Edu \times HPgap$; the instrumental variables are HPR, $FatherIncome \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. 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 is 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.5: Robustness: Impute Household Assets

	(1) Migration	(2) Migration	(3) lnIncome	(4) lnIncome
lnHHAssets× HPgap	0.148***	0.082**	0.372***	0.400***
	(0.030)	(0.033)	(0.127)	(0.136)
HPgap	-1.099***	-0.878**	-2.614	-2.706
	(0.382)	(0.367)	(2.030)	(2.026)
Edu× HPgap		0.016***		-0.007
		(0.003)		(0.015)
lnHHAssets× birthyearFE	Y	Y	Y	Y
K-P F-stat	28.619	19.640	28.126	19.253
LM test	52.366	53.722	52.551	53.811
Obs.	13,980	13,980	11,253	11,253
Mean(Dep. Var.)	0.139	0.139	9.736	9.736

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 $Edu \times HPgap$; the instrumental variables are HPR, $HHAssets \times HPR$, and $Edu \times HPR$. Control variables include gender, parental *Hukou* status, log wage and log employment at the origin prefecture, weighted average of the log wage and log employment at the potential destination prefectures, fixed effects for origin prefecture, birth year, and education level. 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 is 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.6: Conditional Logit: First Stage

	(1) HousingPrices	(2) FatherEdu× HP	(3) HousingPrices	(4) FatherEdu× HP
HPR	-0.137***	-0.415***	-0.121***	-0.407***
	(0.027)	(0.038)	(0.030)	(0.039)
FatherEdu× HPR	0.002	0.693***	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.	603	603	603	603

NOTE: SEs 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.

References

Christoph Albert and Joan Monras. Immigration and spatial equilibrium: the role of expenditures in the country of origin. *American Economic Review*, 112(11):3763–3802, 2022.

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

James Alm, Weizheng Lai, and Xun Li. Housing market regulations and strategic divorce propensity in china. *Journal of Population Economics*, 35(3):1103–1131, 2022.

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

Costas Azariadis and John Stachurski. Poverty traps. *Handbook of economic growth*, 1: 295–384, 2005.

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

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1): 249–275, 2004.

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

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.

Yu Chen and Jufen Wang. Social integration of new-generation migrants in shanghai china. *Habitat International*, 49:419–425, 2015.

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.

Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902, 2016.

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.

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.

Karen Eggleston, Ang Sun, and Zhaoguo Zhan. The impact of rural pensions in china on labor migration. *The World Bank Economic Review*, 32(1):64–84, 2018.

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.

Andrew Foote. The effects of negative house price changes on migration: Evidence across us housing downturns. *Regional Science and Urban Economics*, 60:292–299, 2016.

Qingen Gai, Naijia Guo, Bingjing Li, Qinghua Shi, Xiaodong Zhu, et al. Migration costs, sorting, and the agricultural productivity gap. 2021.

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.

Rufei Guo, Junsen Zhang, and Minghai Zhou. The demography of the great migration in china. *Journal of Development Economics*, 167:103235, 2024.

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

Zhangfeng Jin and Junsen Zhang. Access to local citizenship and internal migration in a developing country: Evidence from a hukou reform in china. *Journal of Comparative Economics*, 51(1):181–215, 2023.

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.

Cynthia Kinnan, Shing-Yi Wang, and Yongxiang Wang. Access to migration for rural households. *American Economic Journal: Applied Economics*, 10(4):79–119, 2018.

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.

Han Li, Jiangyi Li, Yi Lu, and Huihua Xie. Housing wealth and labor supply: Evidence from a regression discontinuity design. *Journal of Public Economics*, 183:104139, 2020.

Lixing Li and Xiaoyu Wu. Housing price and intergenerational co-residence in urban china. *Journal of Housing Economics*, 45:101596, 2019.

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.

Lina Meng, Xiao Xiao, and Yinggang Zhou. Housing boom and household migration decision: new evidence from china. *The Journal of Real Estate Finance and Economics*, 67(3):453–479, 2023.

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

Claudia Olivetti and M Daniele Paserman. In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850–1940. *American Economic Review*, 105(8):2695–2724, 2015.

William W Olney and Owen Thompson. The determinants of declining internal migration. Technical report, National Bureau of Economic Research, 2024.

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

Stephen J Redding and Esteban Rossi-Hansberg. Quantitative spatial economics. *Annual Review of Economics*, 9:21–58, 2017.

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.

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

Hui Ren Tan. A different land of opportunity: The geography of intergenerational mobility in the early twentieth-century united states. *Journal of Labor Economics*, 41(1):77–102, 2023.

Trevor Tombe and Xiaodong Zhu. Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, 109(5):1843–1872, 2019.

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.

Shang-Jin Wei and Xiaobo Zhang. The competitive saving motive: Evidence from rising sex ratios and savings rates in china. *Journal of political Economy*, 119(3): 511–564, 2011.

John Whalley and Shunming Zhang. A numerical simulation analysis of (hukou) labour mobility restrictions in china. *Journal of Development Economics*, 83(2):392–410, 2007.

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

Jeffrey E Zabel. Migration, housing market, and labor market responses to employment shocks. *Journal of Urban Economics*, 72(2-3):267–284, 2012.

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.

Angela Zheng and James Graham. Public education inequality and intergenerational mobility. *American Economic Journal: Macroeconomics*, 14(3):250–282, 2022.