ESSAYS ON INTERCITY MIGRATION IN THE U.S. AND URBAN HOUSING MARKETS IN CHINA

BY

ZHI WANG

B.A., HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY, 2006
M.A., PEKING UNIVERSITY, 2008
M.A., BROWN UNIVERSITY, 2009

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE DEPARTMENT OF ECONOMICS AT BROWN UNIVERSITY

PROVIDENCE, RHODE ISLAND

MAY 2013
This dissertation by Zhi Wang is accepted in its present form by the Department of Economics as satisfying the dissertation requirement for the degree of Doctor of Philosophy

Date___________________          __________________________________
J. Vernon Henderson, Advisor

Recommended to the Graduate Council

Date___________________          __________________________________
J. Vernon Henderson, Reader

Date___________________          __________________________________
Nathaniel Baum-Snow, Reader

Date___________________          __________________________________
Andrew Foster, Reader

Approved by the Graduate Council

Date___________________          __________________________________
Peter M. Weber, Dean of the Graduate School

iii
Curriculum Vitae

Zhi Wang was born on February 7, 1984 in Changsha, Hunan, China and attended Changjun High School. She graduated with honors from the Huazhong University of Science and Technology in 2006 with a Bachelor of Arts in Economics and Peking University in 2008 with a Master of Arts in Economics. She enrolled in Brown University’s doctoral program in Economics in 2008 and obtained her second Master of Arts in Economics in 2009. At Brown, she was a recipient of the Susan R Kamins’ 82 Fellowship and two Graduate Merit Dissertation Fellowships. Prior to beginning graduate work, Zhi additionally studied at Lingnan University, Hong Kong.
Acknowledgements

I first would like to thank Professors Vernon Henderson and Nathaniel Baum-Snow for their help and support over the years. They taught me the core ideas of urban economics and exposed me to various important applied urban topics. Vernon continually and convincingly conveyed a sprint of adventure in regard to research and scholarship. From Nathaniel, I learnt so many useful tools to undertake my own research in urban economics and labor economics. I thank Vernon and Nathaniel for their willingness to contribute both in seminars and in their offices whenever needed. Without their guidance and help this dissertation would not have been possible. I additionally want to thank Professor Andrew Foster. My research has greatly benefited from conversations with Andrew. I also owe much to Qinghua Zhang, who was my mentor in Peking University and has been a wonderful co-author in subsequent years. Lastly, I appreciate the hard work of the economics office staff, especially Angelica Vargas, over the last five years.

I want to thank the Graduate School at Brown University and the Peking University-Lincoln Institute Center for Urban Development and Land Policy for financial support. The Spatial Structures in the Social Sciences at Brown University provided me with a lovely working environment. I also thank many people from the writing center at Brown University for editing my dissertation.

I am indebted to all of my teachers at Fengshushan Elementary School, Changjun Middle and High School as well as my professors at the Huazhong University of Science and
Technology, Lingnan University, Peking University and Brown University. Professors Liutang Gong, Ping Lin and Qinghua Zhang were particularly influential in my development prior to beginning my graduate studies in the United States.

I thank all of my friends and colleagues for their continuous support. Finally, I owe so much to my family. My greatest debt is, as always, to my mother Yu Jin and my father Xiaoling Wang, whose love and support makes everything possible.
Cities are differentiated along many dimensions, with the most important being quality of life and productivity. These attributes guide people’s location choices and migration decisions. Furthermore, this intercity mobility leads the relative attractiveness of a city to be capitalized into its local land and housing prices. In this dissertation, Chapters 1 and 2 contribute to the understanding of the relationship between location decisions and city attributes using individual-level data from the U.S. census. Chapter 3 explores the fundamental explanations for a city’s housing price appreciation and contributes new evidence from urban housing markets in China to the existing literature examining the roles of fundamental factors of supply and demand in real estate markets.

Chapter 1 studies how the location decision on labor market entry affects a worker’s wage growth. Today, especially in the developed world, the most productive knowledge possessed by skilled workers is usually acquired after graduation during the early years of employment. I find that the location where a college graduate first enters the labor force plays an important role in facilitating the acquisition of skills and knowledge and promoting subsequent wage growth. Furthermore, an individual’s learning ability, which is presumably associated with social skills as well as cognitive aptitudes, begets further learning in big cities. Because of that a self-selection by learning ability occurs at labor force entry. I provide empirical evidence that positive sorting by learning ability into big cities when first entering the labor force is a salient feature for college graduates born in
“rural” states in the United States. Estimates imply that in addition to the city-size wage level premium of 15.2 percent, college graduates starting work in big cities experienced 7.2 percent higher wage growth over the five-year period immediately following labor force entry than their counterparts in small cities. Positive sorting by learning ability adds another 3.1 percent to the wage growth difference between big and small cities. While the primary data source for my empirical analysis is the Census Public Use Microdata 5 Percent Sample from 2000, the 1979 National Longitudinal Survey of Youth reveals an experience profile of wages that is consistent with my results from the census.

Chapter 2 extends the intercity migration decision to the later stages of people’s life cycles and proposes a new explanation for falling internal mobility in the U.S. since the 1980s. Following the method in Chapter 1, I measure the evolution of wage growth differentials of American cities over time using the 1980, 1990 and 2000 censuses. Estimates show that the wage growth differentials have been expanding in cities with large population sizes since 1979. By pooling individual-level data from the three censuses, I show that the increase in wage growth gain in relatively large cities has induced people in 1990 and 2000 to stay in these more populated locations longer than their 1980 counterparts. The estimates suggest that, for the 27-35 age cohort in my sample, the evolution of wage growth differentials of cities for labor force entry can alone account for about 74% and 33% of the actual declines in intercity migration rates during the 1980-1990 and 1980-2000 time periods, respectively. By introducing a typical lifecycle migration pattern and examining the impact of the change in city attributes on people’s relocation decisions after labor force entry, this paper improves understanding of the reasons for the trends in internal migration in the U.S. over the past three decades.
Chapter 3 proposes an empirical approach using both city level and residential development project level data to answer two questions: How important are changes in the fundamental factors of supply and demand in explaining housing price appreciation in major Chinese cities? In which cities has price appreciation significantly deviated from changes in fundamentals? Chapter 3 is co-authored with Qinghua Zhang from Peking University.
Contents

1. Smart City: Learning Effects and Labor Force Entry.................................................. 1
  1.1 Introduction............................................................................................................... 1
  1.2 Conceptual framework .............................................................................................. 7
    1.2.1 Baseline model ................................................................................................... 7
    1.2.2 Sorting in the baseline equilibrium................................................................... 10
    1.2.3 Average earnings functions with selection terms............................................. 11
  1.3 Data and Empirical Method .................................................................................... 12
    1.3.1 Data................................................................................................................... 13
    1.3.2 Empirical strategy ............................................................................................. 19
  1.4 Empirical Results .................................................................................................... 23
    1.4.1 The city-specific wage rate and static heterogeneity................................. 23
    1.4.2 The learning effect............................................................................................ 24
    1.4.3 Evidence of positive sorting at labor force entry............................................ 28
    1.4.4 Difference in the distribution of learning ability among birth states............. 31
    1.4.5 City-size cutoff................................................................................................. 33
    1.4.6 Role of natural endowments ............................................................................. 34
  1.5 Evidence from the NLSY79 .................................................................................... 35
  1.6 Conclusions ............................................................................................................. 38
  1.7 Tables ...................................................................................................................... 40
  1.8 Figures ..................................................................................................................... 47
1.9 Appendix A: Selection Thresholds in the Baseline Equilibrium .......................... 54
1.10 Appendix B: Equilibrium Condition for Firms .................................................. 55
1.11 Appendix C: A Distance Model ......................................................................... 56
1.12 Appendix D: A semi-parametric approach ......................................................... 58
1.13 Appendix E: Three city categories ..................................................................... 60

2. Smart City, Life-cycle Migration and Falling Mobility since the 1980s............. 66

2.1 Introduction ........................................................................................................ 66
2.2 Data and Patterns ............................................................................................... 71
  2.2.1 Data ............................................................................................................... 71
  2.2.2 Spatial disparities in wage growth ................................................................. 73
  2.2.3 Migration decisions at and after labor force entry ....................................... 75
  2.2.4 Other individual characteristics .................................................................. 76
2.3 Model .................................................................................................................. 77
2.4 Empirical Results and Discussions ................................................................. 79
  2.4.1 Revolution of wage growth differentials across U.S. Cities, 1980-2000 ...... 79
  2.4.2 Relocation decisions and wage growth differentials of MSAs for labor force entry ................................................................. 82
  2.4.3 Other changes across U.S. cities, 1980-2000 .............................................. 85
  2.4.4 How important is the change in wage growth differentials? ...................... 87
2.5 Conclusions ....................................................................................................... 88
2.6 Tables ................................................................................................................ 90
2.7 Figures ............................................................................................................... 96
2.8 Appendix A: Data Appendix ............................................................................. 104
2.9 Appendix B: Proof of the Proposition ................................................................. 105
2.10 Appendix C: Construct the housing price differentials ................................. 107

3. Fundamentals in China’s Housing Markets ..................................................... 108

3.1 Introduction ..................................................................................................... 108
3.2 Theoretical Framework ................................................................................... 112
3.3 Empirical Strategy ......................................................................................... 114
3.4 Housing Production Function Estimation ....................................................... 115
3.5 Population Elasticity Estimation: Long-difference Specification ................. 120
3.6 Results and Discussions ............................................................................... 127
3.7 Conclusions ................................................................................................... 132
3.8 Tables ............................................................................................................ 133
3.9 Figures .......................................................................................................... 137

3.10 Appendix A: Data for housing production function estimation .................. 139
3.11 Appendix B: Trends of Non-agricultural hukou Population for the 35 Cities and
               Big Events since 1949 ............................................................................... 140
3.12 Appendix C: Definition of cross-province non-hukou migrant workers and the
               2000 census ............................................................................................... 142
3.13 Appendix D: Prediction of cross-province non-hukou migrant workers in 2000 143
3.14 Appendix E: Generate the 95-percent confidence intervals for the calculated
               housing price appreciation rates .................................................................. 146

Bibliography ......................................................................................................... 147
List of Tables

Table 1.1: Summary Statistics ................................................................. 40
Table 1.2: Regression Coefficients and Structural Parameters.................. 41
Table 1.3: Static Heterogeneity of Learning Ability .................................. 42
Table 1.4: Main Results and Parameter Estimates from the Structure Model... 43
Table 1.5: Additional Birth-State Controls .............................................. 44
Table 1.6: Roles of Human and Natural Endowments ............................. 45
Table 1.7: Evidence from the NLSY79 .................................................... 46
Table 1.8: Semi-Parametric Assumption ................................................ 59
Table 2.1: Individual Characteristics, 27-35 .......................................... 90
Table 2.2: Wage Growth Differentials of American Cities, 1980-2000 ......... 91
Table 2.3: Relocation Decisions and Relative Changes in Wage Growth Gain of MSAs for Labor Force Entry ................................................................. 92
Table 2.4: Inclusion of the Changes in Additional Local Price Differentials... 93
Table 2.5: Results under the Alternative Estimates of Wage Growth Differentials..... 94
Table 2.6: Evolutions of Local Price Differentials .................................... 95
Table 3.1: Summary Statistics ................................................................. 133
Table 3.2: Housing Production Function Estimation ............................... 134
Table 3.3: Impact of Population Growth on Housing Price Appreciation ....... 135
Table 3.4: Correlation between the Constructed Hukou Population Growth and Other Local Shocks ................................................................. 136
Table 3.5: Historical non-*hukou* Migrant Flows

............................................................. 145
List of Figures

Figure 1.1: Patterns in the Raw Data ................................................................. 47
Figure 1.2: Share in Big Cities at Labor Force Entry, by Birth State ................. 48
Figure 1.3: Geographic Locations of Birth States and Destinations in the “Big City”
Category .............................................................................................................. 49
Figure 1.4: Positive Sorting at Labor Force Entry—Average of Adjusted Wage Growth 50
Figure 1.5: Positive Sorting at Labor Force Entry—Full Distribution of Adjusted Wage
Growth ..................................................................................................................... 51
Figure 1.6: Parameter Estimates using Different City-Size Cutoffs ....................... 52
Figure 1.7: Experience Profiles of Wages, the NLSY79 ...................................... 53
Figure 1.8: Approximately Linear Relationships between Selected Means and Selection
Probability under Common Probability Distributions ............................................ 58
Figure 2.1: Age Profiles of Interstate and Inter-MSA Migration Rates .................. 96
Figure 2.2: Decomposition of Inter-MSA Migration by Moving Direction .......... 98
Figure 2.3: Average wage growth by Size Category, 1980-2000 ......................... 99
Figure 2.4: Changes in out-Migration Rate against Changes in Wage Growth Differential
by MSA for Labor Force Entry ............................................................................. 100
Figure 2.5: Actual and Predicted Age Profiles of Inter-MSA Migration Rates, 27-35 .. 101
Figure 2.6: Age Profiles of Interstate and Inter-MSA Migration Rates, Subsample ...... 102
Figure 2.7: Decomposition of Inter-MSA Migration by Moving Direction, Subsample 103
Figure 3.1: Trends in Cities’ Average Sale Prices of Residential Housing ............. 137
Figure 3.2: Calculated and Actual Housing Price Appreciation between 2002 and 2008
............................................................................................................................................... 138

Figure 3.3: Trends of Non-agricultural *hukou* Population for the 35 Cities and Big Events since 1949 .............................................................................................................................................. 141
1. Smart City: Learning Effects and Labor Force Entry

1.1 Introduction

Cities are differentiated by the extent to which they promote wage growth as well as by the level to which they enhance productivity. A vast literature on agglomeration economies provides considerable empirical evidence of productivity level differences: workers gain an immediate wage level premium by moving to areas with large or densely concentrated populations.\(^1\) Within this literature, only a few empirical studies consider spatial disparities in wage growth; these studies show that in bigger cities, the return to experience is higher than in small cities, and the relevant human capital gain appears to be portable for workers who then move to other cities.\(^2\) This growth attribute, called the city’s *learning effect*, affects wage growth by influencing the acquisition of skills and knowledge for young workers.\(^3\) Therefore, the decision to live in a big city encompasses not only the choice of residence, but also investment in human capital in early career.


\(^{3}\) Previous studies consider two potential mechanisms for urban wage growth premium: faster learning and better job/task matches in bigger cities. Yet it would be hard to empirically separate one from the other. Better matching quality in bigger cities may result from either more efficient within-city, between-job transfers (e.g., thicker labor markets) or better within-firm tasks assignments (e.g., better management skills of employers). Therefore, for a young worker, the process of matching himself to the best job/task would usually be accompanied by the gain of knowledge about his comparative advantage. I model this growth attribute as the learning effect (human capital accumulation), but bear in mind that better coordination of labor markets could be another explanation.
For first-time job seekers, big cities are attractive destinations because they enhance productivity and facilitate learning, but this attractiveness is likely to be offset by high costs of living. Furthermore, young workers with higher unobserved ability may acquire more knowledge and skills, resulting in wage growth dispersion among people with heterogeneous learning ability. In this case, a non-random selection would occur at labor force entry if learning ability is more highly rewarded in bigger cities. In the United States, about 60 percent of college graduates changed their metropolitan areas (MSAs) of residence when they first entered the labor force. Despite the significance of this pattern, little work has been done on how the location decision on labor market entry shapes subsequent wage growth. This paper investigates the properties of the learning effect and its role in location decisions on labor force entry of people endowed with heterogeneous learning ability. To pursue this question, I estimate model parameters that characterize the learning effect with a large sample of college graduates who entered the labor force in nearly 300 U.S. cities. The results indicate that, besides higher entry wage level, wage growth is more rapid in bigger cities, especially for young workers with high learning ability. However, such return to learning ability is absent in small cities and rural areas. Estimates imply that in addition to the city-size wage level premium of 15.2 percent, college graduates who entered the labor force in big cities experience 7.2 percent higher wage growth over the five-year period immediately following labor force entry than their counterparts in small cities. Positive sorting by learning ability adds another 3.1 percent to the wage-growth gap between big and small cities.

4 Theory and empirics in urban economics reckon the existence of heterogeneity among observationally identical individuals, and the unobserved ability tends to be better rewarded in bigger cities in the form of a larger return to experience (Gould, 2007; Baum-Snow and Pavan, 2011; Davis and Dingel, 2012; De la Roca and Puga, 2012).
For the empirical analysis, I develop a simple multiple-city, two-period model. The model characterizes the learning effect and incorporates this characterization into a framework similar to Roback (1982) to study people’s location decisions on labor force entry. The learning effect of the city where an individual enters the labor market affects the individual’s wage growth by influencing human capital accumulation. Individuals are endowed with unobserved learning ability. By affecting the amount of a worker’s human capital, learning ability could generate both static and dynamic heterogeneities in his wages. Between the two periods in my model, the individual decides whether to switch cities. Human capital accumulated in period one is portable when a worker moves to another city. In equilibrium, individuals with higher learning ability enter the labor force in cities where this ability is better rewarded.

As the primary data source for my empirical analysis, the Census Public Use Microdata 5 Percent Sample from 2000 provides a sufficiently large sample of cities in the contiguous United States. More importantly, using a variety of birth states and a sufficiently large sample of people from each birth state, I construct variation in average learning ability across birth states to address the issue that learning ability is generally unobserved. Within city-size categories, diverse and large samples are essential for identifying the rate of return to learning ability in each category. I use metropolitan statistical areas (MSAs) to define cities. I focus on white males with a bachelor’s degree.

To estimate parameters that characterize the learning effect, I use two adjacent age cohorts — 22-26 and 27-31 — from the 2000 census as the empirical counterparts for the model. I assume that the distribution of learning ability remains unchanged between the
cohorts, and that all city characteristics are constant between 1995 and 2000. I identify location choices of these two cohorts using their mobility histories. To measure wage growth, I estimate the city-specific wage rate from the earnings of the younger cohort, which I use as a proxy for the entry wage rate of the older cohort. The general patterns in the raw data suggest that college graduates entering the labor force in big cities have, on average, 7.9 percent higher wage growth than in small cities, and they do not make up a random sample of the population. The geographic locations of birth states relative to available destinations in the contiguous U.S. create variation in average learning ability within a destination across birth states. Using this birth-state-level variation and applying a parametric assumption to a structural model of location choices whether to enter the labor force in a big city versus a small city, I estimate the city’s average learning effect and the rates of returns to learning ability in big cities as well as small cities and rural areas. I assume in the baseline estimation that recent college graduates born in different birth states share a single learning ability distribution. Moreover, I explicitly incorporate the possible difference in the distribution of learning ability among birth states into the estimation specification. This possible difference is likely to cause me to underestimate the rate of return to learning ability in big cities. Additionally, the 1979 National Longitudinal Survey of Youth (the NLSY79) reveals an experience profile of wages that is consistent with my main results from the census.

This paper contributes to the existing literature in several areas. First, this paper builds upon the findings of recent empirical studies that investigate reasons for urban wage premium and find higher wage growth in larger cities (Glaeser and Mare, 2001; Gould, 2007; Baum-Snow and Pavan, 2011; De la Roca and Puga, 2012). Instead of tracking
wage profiles with panel data, I measure wage growth using the census. This method allows me to generate evidence of wage growth differentials across nearly 300 U.S. cities. Based on this large sample of cities for labor force entry, I explore what kind of city characteristics are associated with this growth attribute beyond population size, which is the only dimension emphasized in previous empirical work. Second, there is no consensus among previous studies on the role of sorting on unobserved ability in the city-size wage premium. Combes et al. (2008) claim that individual skills account for a large fraction of existing spatial wage disparities. Other studies suggest that the urban wage premium is not the result of omitted ability variables (Glaeser and Mare, 2001; Baum-Snow and Pavan, 2011; De la Roca and Puga, 2012). Individual heterogeneity cannot be characterized as the individual-fixed effect with panel data when the rate of return to unobserved ability varies by destination. To resolve this issue, I use variation in average learning ability within city size categories across birth states to identify the rate of return to learning ability in each city size category. Based on the estimates for model parameters, I am able to provide clear evidence of positive sorting by learning ability into big cities at labor force entry. Furthermore, this paper claims that sorting on unobserved learning ability contributes to the observed city-size wage premium by inflating the city-size wage growth difference. In addition, previous studies find that pecuniary returns generally accumulate over a multiple-year period following interstate migration (Shaw, 1991; Borjas, Bronars, and Trejo, 1992). I present similar patterns: among workers entering the labor market in big cities, people born in states far away from those big cities, on average, have greater wage growth than people originally from states in which those big cities are located. Besides the notion that migrants invest heavily into a new stock of location-
specific capital following migration, I suggest that self-selection on unobserved ability provides another explanation for the differential in pecuniary returns between natives and migrants. Third, the literature on the “smart city” focuses on the role of a city’s human capital stock in enhancing the growth in local productivity and quality of life to explain the positive correlation between a city’s human capital stock and the city’s employment growth (Glaeser and Saiz, 2003; Shapiro, 2006). My results suggest that the “smart city” facilitates human capital accumulation for young workers, thereby attracting first-time job seekers. Fourth, this paper complements studies that evaluate the values of cities (Roback, 1982; Albouy, 2008), by suggesting that a city’s learning effect is also capitalized into the city’s cost of living, along with quality of life and local productivity. Lastly, this paper is relevant to studies on the impact of adverse labor market entry conditions on lifetime labor market performance (Oyer, 2008; Oreopoulos, et al., 2008; Kahn, 2010; Brunner and Kuhn, 2010). In addition to the fluctuation of business cycles as the source of variation used in previous research, this paper demonstrates that labor market entry conditions also depend on people’s birth states. Given the actual destinations at which recent college graduates entered the labor force, the average wage growth for people born in states like DC, NJ, NY, MD and MA is about 6 to 7 percent higher than in states like IA, SD, AR, WV and MT.

The rest of the paper is structured as follows. Section 1.2 presents the theoretical framework. Section 1.3 discusses the data and the empirical method. Section 1.4 presents the estimation results and discussions. Section 5 provides the evidence from the NLSY79. Finally, Section 1.6 concludes.
1.2 Conceptual framework

This section presents a simple model to motivate the empirical analysis. This model builds on a standard inter-city spatial equilibrium framework (Roback, 1982) and incorporates the learning effect. Rather than a full life-cycle model, I use a two-period model since my data only allows me to track where one first enters the labor force for two age groups. For the purpose of estimation, I specify the function of average earnings in each period, conditional on location choices.

1.2.1 Baseline model

The national economy is closed and contains $J$ cities, denoted by $j = 0, \ldots, J-1$. Cities differ in three attributes: 1) the quality-of-life ($Q$), which raises an individual’s utility; 2) the local-productivity in the traded-good sector ($A$); and 3) the learning effect ($g$), which influences human capital accumulation. All city attributes depend on a vector of city characteristics.\(^5\)

There is a continuum of fully mobile individuals in the economy. They are identical except for learning ability $a$, $a \in (-\infty, +\infty)$, which generates both static and dynamic heterogeneities in wages by affecting the amount of human capital as specified below. Learning ability is associated with both cognitive and social skills prior to labor force entry. Each individual (worker/consumer) has two periods in his working life, and he always knows his learning ability. In each period, the individual consumes the numeraire

\(^5\) The city characteristics are either exogenously given (e.g., climate, geology, geography, etc.) or endogenously determined but stationary over time (e.g., MSA population size).
traded good and inelastically demands a single unit of land that he rents at market rate \( p_j \).

The flow payoff is specified as follows:

\[
U_t = \phi_1 \log w_t + \phi_2 \log Q - \phi_3 \log p, \quad t = 1, 2,
\]

where \( w_t \) represents an individual’s wage in period \( t, t = 1, 2 \), \( p \) is the land rent, and \( \phi_1, \phi_2, \) and \( \phi_3 \) are positive.\(^6\)

For the individual endowed with \( a \), in period one the human capital stock is \( \exp(\mu \times a) \), where \( \mu \geq 0 \) (static heterogeneity); in period two the human capital stock is \( \exp(\mu \times a + g_j(a)) \), where \( g_j(a) \) represents city \( j \)’s learning effect, and \( j \) indicates the city where the individual works in period one. A city’s learning effect influences the amount of human capital accumulated in period one by affecting the acquisition of skills and knowledge. The extent of the city’s learning effect has the following function form:

\[
g_j(a) = g_j + \pi_j \times a, \quad j = 0, \ldots, J - 1, \quad \text{where } \pi_j \geq 0 \text{ (dynamic heterogeneity)}. \]

\( g_j \) is the city’s average learning effect. \( \pi_j \times a \) represents the individual-specific learning effect the individual endowed with a higher learning ability is more able to acquire knowledge and skills. Furthermore, learning ability is rewarded differently across cities in terms of the efficiency of learning, and the ranking of the rate of dynamic return to learning ability, \( \pi_j \), is given by: \( \pi_{j-1} > \pi_{j-2} > \ldots > \pi_1 > \pi_0 \).

---

\(^6\) Land represents any location-fixed good. Therefore, the differentials in \( p \) across localities stand for the differentials in the cost of living.
Identical and fully-mobile firms produce the numeraire traded good. Labor $L$ and capital $K$ enter as factors of production. Let $F(K, L; A_j)$ be the production function in city $j$, with a constant-returns-to-scale technology. $L$ is measured by a worker’s human capital stock. In equilibrium a city’s productivity ($A_j$) determines the city’s city-specific wage rate, denoted by $\log w_j$. Individual labor supply is inelastic in both periods. Therefore, in each period, an individual’s earnings depend on both the city-specific wage rate in the city of current residence and his human capital stock. For the individual endowed with $a$ who lives in city $j$ in period one and in city $k$ in period two, the logarithm of earnings in period one and in period two, $\log w_{j,t=1}$ and $\log w_{j,t=2}$, are given by:

\[(2a) \log w_{j,t=1} = \log w_j + \mu \times a\]

\[(2b) \log w_{j,t=2} = \log w_k + \mu \times a + g_j(a) + \epsilon_{jk}\]

where $j, k \in \{0, 1, ..., J - 1\}$, and $\epsilon_{jk}$ is a stochastic term characterizing the quality shock of the job offer arriving in city $j$ from city $k$ at the end of period one. $\epsilon_{jk}$ affects the decision of switching cities between periods and is independent over $j$, over $k$, and over $a$. Any positive correlation between learning ability and the quality of the second-period job offer would be already included in the individual-specific terms in (2b).

The sequence of an individual’s location choices is as follows. Prior to period one, the individual decides where first to enter the labor force, based on the expected lifetime utilities of potential destinations. By the end of period one, the individual decides
whether to switch cities. The individual observes equilibrium local prices $w_j, p_j$, the
quality-of-life $Q_j$, as well as $g_j, \pi_j$, and $\mu$. They are all stationary over time.

1.2.2 Sorting in the baseline equilibrium

In equilibrium each individual makes his migration decisions so as to maximize his
expected lifetime utility, and there exist $J-1$ selection thresholds

$$a_j^* = -\frac{(\Psi_{j+1} - \Psi_j) + \beta \phi_j(g_{j+1} - g_j)}{\beta \phi_j(\pi_{j+1} - \pi_j)}$$

where $\Psi_j = \phi_1 \log w_j + \phi_2 \log Q_j - \phi_3 \log p_j$, $j = 0, \ldots, J-2$, and $\beta \in (0,1)$ is the discount
rate, such that at labor force entry the individual with $a \in (a_{j-1}, a_j^*)$ chooses city $j$, $j = 1, \ldots, J-2$; the individual with $a \in (-\infty, a_0^*)$ chooses city 0; the individual with $a \in (a_{j-2}^*, +\infty)$ chooses city $J-1$. $\Psi_j$ corresponds to the first-period utility for the
average individual ($a = 0$) in city $j$. See Appendix A for details. In equilibrium, individuals partition themselves into cities based on the ranking of $\pi_j$ and their learning
abilities. For example, the individual with learning ability above $a_{j-1}^*$ but below $a_j^*$ finds
entering the labor force in city $j$ generates the highest expected lifetime utility than in
the other cities; the individual with $a_j^*$ is indifferent between city $j$ and city $j+1$ for labor force entry.
Another implication of the baseline equilibrium is as follows. Individuals endowed with $a_j^*, j = 0, ..., J - 2$, are marginal persons, who drive the differential in the local cost of living (i.e., differences in $p_j$) between any two cities in equilibrium:

$$
\phi_j(\log p_j - \log p_k) = \phi_1/\varphi_L(\log A_j - \log A_k) + \phi_2(\log Q_j - \log Q_k) + \beta \phi_3(g_j - g_k) + \sum_{h=k}^{J-1} a_h^* \beta \phi_4(\pi_{h+1} - \pi_h)
$$

where $\varphi_L$ is firm’s expenditure share on labor. The equation above implies that a city’s average learning effect is capitalized into the city’s cost of living, along with the other two traditional city attributes, quality-of-life and the local-productivity. The cost of investing human capital by entering the labor force in cities with greater average learning effects is paid by higher costs of living.

1.2.3 Average earnings functions with selection terms

Based on (2a) and (2b) and the selection rule in equilibrium, I specify the function of average earnings in each period, conditional on location choices. Given $j$ as the location in period one, the average wage rate in period one is:

$$
E(\log w_{j,t=1} | j) = \log w_j + \left\{ \begin{array}{ll}
\mu E(a | a > a_{j-2}^*) & \text{if } j = J - 1 \\
\mu E(a | a_{j-1}^* < a < a_j^*) & \text{if } j = 1, ..., J - 2 \\
\mu E(a | a < a_0^*) & \text{if } j = 0
\end{array} \right.
$$

Given $j$ as the location in period one and $k$ as the location in period two, the average wage rate in period two is

---

7 Appendix B illustrates the equilibrium condition for firms.
where \( E(a \mid a > a^{*}_{j-2}) \), \( E(a \mid a^{*}_{j-1} < a < a^{*}_{j}) \), and \( E(a \mid a < a^{*}_{0}) \) are selected means of learning ability. A city’s average learning effect (\( g_{j} \)) and the city’s rate of dynamic return to learning ability (\( \pi_{j} \)) characterize the city’s learning effect. In the empirical analysis, I estimate \( g_{j} \) and \( \pi_{j} \).

### 1.3 Data and Empirical Method

For the individual endowed with learning ability \( a \), city \( j \)’s learning effect is given by \( g_{j}(a) = g_{j} + \pi_{j} \times a \). I propose an empirical method to estimate \( g_{j} \) and \( \pi_{j} \) using a large sample of college graduates who entered the labor force in nearly 300 U.S. cities. I use information on both earnings and migration histories from the 2000 census for empirical analysis.

To identify the rates of returns to learning ability in each destination (\( \mu \) and \( \pi_{j} \)), it is necessary to have certain variation of selected means in equation (3a) and (3b) across identifiable groups of individuals. The geographic locations of birth states relative to available destinations in the contiguous United States create this variation. This method requires that individuals in each destination at labor force entry are drawn from a variety of birth states in terms of the relative geographic locations. Furthermore, this method requires that in each destination there are a sufficient number of individuals from each
birth state. Another underlying estimation requirement is that the ranking of $\pi_j$ is known to assign the corresponding selected mean to each city. I thus group cities by population size and rank them based on a prior belief. Categorizing destinations into “small city” versus “big city” categories is a natural starting point, since it is the least arbitrary way of partitioning cities according to the prior belief on the ranking of their $\pi_j$. I use 1.5 million as the city-size cutoff and experiment with the other values. Since the existing empirical evidence suggests that the unobserved ability is better rewarded in bigger cities in the form of a larger return to experience (Gould, 2007; Baum-Snow and Pavan, 2011; De la Roca and Puga, 2012), I assume that $\pi_j$ in big cities is higher than in small cities and rural areas.

Although my main results shown in Section 1.4 are based on the case of two city-size categories, in general, my method can be applied to the scenario where there exists multiple city categories, as long as the ranking of $\pi_j$ is known, and the sample size requirements are satisfied. In Appendix E, I extend the empirical analysis to a case of three city categories and allow for more dimensions based on which cities are grouped.

1.3.1 Data

The primary data source for analysis is the Census Public Use Microdata 5 Percent Sample from 2000 (Ruggles, et al., 2010). The census has the following advantages for my empirical implementation. First, it provides a large sample of MSAs in the contiguous United States as destinations for labor force entry. Second, in each city-size category, a large number of observations from a variety of birth states are essential for capturing the
variation of selection terms. I construct information on employment, wages, and mobility history (current residence and residence five years ago) for a sample of white men, aged 22-31 on December 31, 1999, with a bachelor’s degree, whose locations at labor force entry are observed. Age groups 22-26 and 27-31 correspond to period-one and period-two individuals, respectively. I limit my analysis to those who reported working at least 40 weeks, 35 usual hours per week and who earned at least 75 percent of the federal minimum wage in 1999. My earnings measure is the log hourly wage calculated as the logarithm of wage and salary income divided by the product of weeks worked and usual hours worked per week. The full-time full-year limitation allows me to focus on individuals who are less likely to be constrained in their residential locations by family or education considerations. I use white men only to alleviate the concerns of discrimination and labor market attachment for women and non-whites. I also exclude observations for those who worked in sectors of agriculture, fishing, forestry, hunting, and mining, who reported military history, or who were born outside of the contiguous United States. I use MSAs to define cities. In my sample, I have 327 localities as destinations for labor force entry in the contiguous United States. Among them, 281 are MSAs, for which I have MSA-level variables. Another 46 are treated as rural areas in each state. The MSA-level variables such as population size and the share of college graduates in the labor force are calculated as aggregates from the 2000 census. An MSA’s population density is derived by dividing the population size by the land area.

---

8 For cohort 22-26, MSAs of current residence are assumed to be locations where they first enter the labor markets; for cohort 27-31, their MSAs of residence five years ago are assumed to be locations where they first enter the labor force.

9 As Baum-Snow and Pavan (2011) suggest, measurement error is an additional justification for using full-time full-year workers. Baum-Snow and Neal (2009) demonstrate that there exits significant measurement error in hourly wages for part-time and part-year workers in the census.

10 I also exclude individuals who report Hawaii or Alaska as his state of birth.
which is provided by the U.S. Census Bureau. An MSA is classified into the “big city” category if its population size is over 1.5 million. An MSA of less than 1.5 million people and rural areas (non-MSAs or unidentified localities in an identified state) are classified into the “small city” category. I have 36 MSAs in the “big city” category (big cities) as destinations for labor force entry in my sample.

Table 1.1 presents summary statistics for two age cohorts. For both age groups, nearly half of them entered the labor force in big cities. In age group 27-31, 43.7 percent switched localities during the five-year period after labor force entry. Across age cohorts, weeks worked in 1999 and usual hours worked per week in 1999 are roughly the same, while there is a 34 percent growth in the hourly wage. The distribution across industry sectors remains stable across age groups.

To quantify the learning effect, I construct a wage growth measure using two adjacent age cohorts, 22-26 and 27-31. The hourly wage rate of the older cohort, \( \log w_{jk, t=2} \), consists of two components: the wage level component from the MSA of current residence \( k \) (five years after labor force entry) and the wage growth component from the MSA \( j \) for labor force entry. I treat the city-specific wage rate in the city of current residence estimated from the earnings of the younger cohort (\( \log w_k \)) as a proxy for the wage level component in the hourly wage rate of the older cohort, assuming that ability distribution is fixed between two age cohorts, and that differentials in all three city attributes remain constant between 1995 and 2000.\(^{11}\) Among the older cohort, the wage

---

\(^{11}\) Chen and Rosenthal (2008) estimate the quality of life and local productivity of American cities for each decade from 1970 to 2000. They find that the correlation between 1990 and 2000 for the perceived quality
growth of individual $i$ who worked in MSA $j$ when first entering the labor market and currently lives in MSA $k$, $\Delta \log w_{ijk}$, is calculated as follows:

$$\Delta \log w_{ijk} = \log w_{ijk,t=2} - \log w_k$$

Intuitively, after controlling for the wage level differences in cities of current residence, variation in the earnings of the second period captures the wage growth differentials of cities at labor force entry. Mathematically, Equation (2b) suggests that if there is no static heterogeneity (i.e., $\mu = 0$), $\Delta \log w_{jk}$ measures exactly the learning effect from city $j$ (plus a stochastic term); if there exists static heterogeneity, $\Delta \log w_{jk}$ would also include the static heterogeneity generated in the city of current residence (city $k$). As discussed later, my empirical results suggest that the static heterogeneity is negligible. Panel A in Figure 1.1 graphs the average wage growth within each city-size category. (I standardize $\Delta \log w_{jk}$ so that individuals in the “small city” category have zero average wage growth.) On average, entering the labor force in cities of the “big city” category is associated with a significant 7.9 percent wage growth premium.

Another important feature in the raw data is that the share of college graduates in big cities at labor force entry varies largely by people’s birth states. Figure 1.2 presents the shares across 49 states of birth (including DC and excluding Hawaii and Alaska), calculated from both cohorts. For each age group, the range of this share goes from about 0.14 to 0.80; the distribution stays roughly stable across two cohorts. An obvious pattern of life and business environment are 88.07 percent and 93.21 percent, respectively, indicating that there is considerable persistence in these two city attributes associated with a given location.
is that recent college graduates born in “rural” states (states that do not have a big city located within their boundaries), such as AR, LA, MS, AL, TN, KY, are less likely to start their first jobs in big cities. The majority of recent college graduates born in “urban” states (states in which big cities are located), such as DC, MD, IL, AZ, CA, enter the labor force in big cities. Figure 2.3 maps MSAs of more than 1.5 million people in 2000 in the contiguous United States (blue circles). States in dark green are “rural” states and states in light green are “urban” states. The geographic locations of hometowns relative to available destinations seem to exert great influence on location decisions at labor force entry. Given the prior belief on the ranking of $\pi_j$ (i.e., learning ability has a higher return in big cities than in small cities and rural areas), if the distance between an individual’s hometown and a destination lowers the attractiveness of this destination, the marginal individual (who is indifferent between starting work in a big and a small city and whose learning ability pins down the selection threshold) born in a “rural” state needs to have a relatively high learning ability to offset the effect of distance. This variation in selection threshold across birth states would lead to a variation in average learning ability within each city-size category across birth states. Specifically, after entering the labor force and within a city-size category, young workers born in “rural” states would have higher average learning ability than young workers born in “urban” states, thereby leading to greater average wage growth. Panels B and C in Figure 1.1 depict the consistent pattern from the raw data. Panel B shows that average wage growth is higher for people born in “urban” states since they are more likely to work in big cities in early career. However,

---

12 In the contiguous United States, 22 of 49 states have an MSA of more than 1.5 million people in 2000. An MSA’s boundaries may cross states’ boundaries. For example, NJ and PA share the Philadelphia metropolitan area, and MO and IL share the St. Louis metropolitan area.
when further dividing workers by the city size of destinations at labor force entry, Panel C shows that workers born in “rural” states have significantly higher average wage growth than workers from “urban” states in the “big city” category. However, such wage growth difference between “rural” and “urban” states is insignificant within the “small” city category, suggesting a negligible role of dynamic return to learning ability in small cities and rural areas. (In Panels B and C, I standardize $\Delta \log w$ so that within the “small city” category individuals born in “urban” states have zero average wage growth.)

Analogous to Panel B, Panel D shows that average entry wage rate is higher for people born in “urban” states. Panel E graphs this average entry wage rate for each city-size category at labor force entry, repeated for ”urban” and “rural” states. (In Panels D and E, I standardize the entry wage rate so that within the “small city” category workers born in “rural” states have zero average entry wage rate.) Compared with Panel C, in Panel E of Figure 2.1, the difference in average entry wage rate within each city-size category is insignificant between ”rural” and “urban” states, suggesting that the static heterogeneity plays a negligible role.

---

13 One concern is that the major destinations in the “big city” category differ between ”rural” and “urban” states, and therefore the difference in average wage growth may be caused by the variation in the city’s average learning effect across destinations. The data suggest that college graduates from these two types of states share the same major destinations in the “big city” category (i.e., Atlanta, Chicago, DC, Boston, Dallas, New York, Los Angeles, and Minneapolis). Additionally, I plot a similar graph controlling for city-fixed effects of destinations at labor force entry. The pattern in Panel C still persists. Another concern is that some destinations in the “big city” category for “rural” states in the south may have a higher $\pi_j$. Evidence from the case of three city category suggest that even if there exists a difference in $\pi_j$ within the “big city” category, big cities in the south (e.g., Dallas, Houston, Phoenix, and Orlando) are more likely to have a lower $\pi_j$ than big cities such as New York, Boston, Washington, DC, Chicago, and Minneapolis.
1.3.2 Empirical strategy

I categorize cities (MSAs, non-MSAs, and unidentified localities in an identified state) into two categories based on population sizes. I assume that cities in each category share the same $\pi_j$. $\pi_i$ and $\pi_0$ denote the rate of dynamic return to learning ability for the “big city” and the “small city” categories, respectively. The prior belief on the ranking of $\pi_j$ is: $\pi_i > \pi_0$.

In the reality, individuals are not fully mobile, and therefore distances matter in migration decisions. The migration literature suggests that people value social connections to the place of origin (e.g., family, friends, and other social networks), and distances weaken these types of connections. Therefore, working in a destination far away from one’s family causes certain kinds of psychic costs. Additionally, as Schwartz (1973) suggests, these kinds of psychic costs may then be transformed into permanent transportation costs due to the needed visits to the place of origin so as to negate the agony of departure from family and friends. A distance model presented in Appendix C shows that unequal geographic distances create the birth-state-specific selection threshold, causing the average learning ability within each city-size category to vary by birth state. Using this variation in average learning ability within each city-size category across birth states, I estimate the rates of returns to learning ability in this category (i.e., $\mu$, $\pi_0$, and $\pi_i$).
To specify the equations of estimation based on (3a) and (3b), I assume that learning ability is normally distributed with mean at zero and variance at $\sigma^2_a$. I thus rewrite the birth-state-specific selected means of learning ability as follows:

$$
E(a \mid a > a^*_S) = \sigma_a \lambda_1(a^*_S / \sigma_a), \quad \text{and} \quad E(a \mid a < a^*_S) = -\sigma_a \lambda_0(a^*_S / \sigma_a),
$$

where $\lambda_1(c) = \phi(c) / (1 - \Phi(c))$, and $\lambda_0(c) = \phi(c) / \Phi(c)$. $\phi$ represents the PDF of standard normal distribution, and $\Phi$ represents the CDF of standard normal distribution.

Furthermore, the share of college graduates from birth state $S$ who enter the labor force in big cities, denoted as $P_S$, is given by: $P_S = 1 - \Phi(a^*_S / \sigma_a)$. Based on this relationship between the unobserved selection threshold $a^*_S$ and the observed selection probability $P_S$, I write the selected means of learning ability as functions of $P_S$. For my sample, both $\lambda_1$ and $\lambda_0$ are approximately linear in $P_S$. It suggests that the problem of multicollinearity would occur if I include both estimated $\lambda_1$ and $\lambda_0$ in the regression. To resolve this issue, I impose the following linear approximations on $\lambda_1(P_S)$ and $\lambda_0(P_S)$ in the regressions: $\lambda_j(\Phi^{-1}(1 - P_S)) = \hat{\lambda}_j + \delta_j \times P_S$, $j = 0, 1$. Given this assumption, within each city-size category the average learning ability of people born in state $S$ linearly decreases with $P_S$. For my empirical analysis, I use $P_S$ calculated from the younger cohort and place $\delta_0$ at 1.582 and $\delta_1$ at -1.910.

To construct the wage growth measure, I estimate the city-specific wage rate, log $w_k$, in the following log earnings equation for the younger cohort based on (3a):

---

14 Note that the normality assumption is not strictly necessary for my general implications as discussed in Appendix D.
(4) $\log w_{ik,t=1} = \tau_k + \phi_1 \times T_i \times P_s + \phi_2 \times P_s + \tilde{X}_i \cdot \tilde{\eta} + u_{ik,t=1}$

where $\log w_{ik,t=1}$ is the log hourly wage for individual $i$ who works in city $k$, $\phi_1 = (\mu \sigma_\delta \delta_1 + \mu \sigma_\delta \delta_0)\cdot$; $\phi_2 = -\mu \sigma_\delta \delta_0\cdot$; $\tilde{X}_i$ includes age, age squared, marital status, and industry dummies, $T_i$ indicates whether individual $i$ enters the labor force in a big city, and $u_{ik,t=1}$ is the transitory component of log earnings in period one. The city-specific wage rate, $\log w_k$, is estimated from the MSA fixed effect $\tau_k$. Because the main effect of $T_i$, which is $\mu \sigma_\delta (\tilde{\lambda}_1 + \tilde{\lambda}_0)$, is subsumed by MSA dummies, I can obtain consistent estimates of $\log w_k$ only if there exists no static heterogeneity (i.e., $\mu = 0$). My empirical results show that $\mu$ is negligible.

For the purpose of estimating the parameters that characterize the learning effect ($g_j$, $\pi_1$, and $\pi_0$), I calculate the wage growth measure using the earnings of the older cohorts and the estimate of the city-specific wage rate: $\Delta \log w_{ijk} = \log w_{ijk,t=2} - \hat{\log w_k}$, where $\log w_{ijk,t=2}$ is the log hourly wage of individual $i$ in city $k$ who worked in city $j$ five years ago, and $\hat{\log w_k}$ is the estimate of the city-specific wage rate in city $k$. Human capital accumulation is more efficient in bigger and denser cities, and young workers learn more efficiently when surrounded by the other highly-educated workers (Glaeser, 1999; Peri, 2001; Moretti, 2004; Davis and Dingel, 2012), suggesting that the related city characteristics (i.e., city size, population density, and human capital stock) could

---

15 I treat non-MSA areas and unidentified localities in one state as a single rural area. MSA dummies in equation (4) include 292 MSAs and 47 rural areas.
influence the efficiency of acquiring skills and knowledge and thereby the level of the city’s average learning effect (\( g_j \)). Therefore, I impose the following relationship between \( g_j \) and the relevant city characteristics: 
\[
g_j = \gamma_1 Z_{1j} + \gamma_2 Z_{2j} + \gamma_3 Z_{3j},
\]
where \( Z_{1j} \) represents the log population in MSA \( j \), \( Z_{2j} \) is the share of college graduates in the labor force in MSA \( j \), and \( Z_{3j} \) is the population density of MSA \( j \). Based on (3b), I estimate the following equation of the wage growth, \( \Delta \log w_{ijk} \), for the older cohort:

\[
(5) \quad \Delta \log w_{ijk} = \gamma_1 Z_{1j} + \gamma_2 Z_{2j} + \gamma_3 Z_{3j} + \alpha_1 \times T_i + \alpha_2 \times T_i \times P_s + \alpha_3 \times P_s + X_i \eta + u_{ijk,t=2}
\]

where 
\[
\alpha_1 = (\mu + \pi_1)\sigma_a \tilde{\lambda}_1 + (\mu + \pi_0)\sigma_a \tilde{\lambda}_0,
\]
\[
\alpha_2 = (\mu + \pi_1)\sigma_a \tilde{\delta}_1 + (\mu + \pi_0)\sigma_a \tilde{\delta}_0,
\]
\[
\alpha_3 = -(\mu + \pi_0)\sigma_a \tilde{\delta}_0,
\]
\( X_i \) includes age, age squared, marital status, industry dummies, and a constant term, and \( u_{ijk,t=2} \) is the transitory component of log earnings in period two.

Equation (5) is the main regression specification for estimating parameters from the structural model. From the regression of (5), I directly estimate \( \gamma_1 \), \( \gamma_2 \), and \( \gamma_3 \). Based on their estimates, I calculate the city’s average learning effect for 281 MSAs in the sample. Panel A in Table 1.2 presents the key coefficients in specifications 4 and 5, along with their relationships with parameters from the model, based on which I infer \( \mu \sigma_a \), \( \pi_0 \sigma_a \), and \( \pi_1 \sigma_a \).

An econometric issue is that within an MSA (where one entered the labor force) and birth state the main regressors in specifications (4) and (5) (except for the individual

\[\text{16} \quad \text{Since } \sigma_a \text{ is unknown (the learning ability does not have a well-defined unit of measurement), } \pi_1 \text{ and } \pi_0 \text{ can only be estimated up to scale.}\]
characteristics) do not vary by worker. Therefore, I face the classical Moulton (1990) problem of estimating the effects of aggregate variables on individual outcomes. I deal with this problem by having the standard errors clustered at the birth state/MSA level. Additionally, all tables and graphs in this paper are based on appropriately weighted data.

1.4 Empirical Results

1.4.1 The city-specific wage rate and static heterogeneity

Panel A of Table 1.3 reports the regression estimates of $\phi_1$ and $\phi_2$ in specification (4) for both the full sample of the younger cohort (column 1) and a subsample excluding individuals entering the labor force in a non-MSA or an unidentified locality (column 2). In both regressions, these two coefficients are insignificant and have small magnitudes. Panel B of Table 1.3 presents the estimates of $\mu \sigma_\delta \delta_1$ and $\mu \sigma_\delta \delta_0$. Given that $\delta_0$ and $\delta_1$ are non-zero numbers, the estimates in Panel B suggest that the rate of static return to learning ability, $\mu$, is zero. For the full sample the F-statistic for testing the null hypothesis that $\phi_1$ and $\phi_2$ are zero is 0.15 with the p-value at 0.86, and for the subsample the F-statistic is 0.11 with the p-value at 0.89, further suggesting that $\mu = 0$. This result implies that for the college graduates in my sample, learning ability does not affect wages before they start to acquire skills and knowledge following labor force entry. For the

---

17 Because $P_s$ is calculated instead of being directly observed, inference based on the estimates from regressions of specifications (4) and (5) may suffer from the problem of generated regressor. To check the robustness of inference, I also calculate the standard errors and test statistics on the basis of 1000 bootstrap replications. The results of all the tests remain unchanged.

18 The MSAs of current residence for 6265 of the 35063 individuals in the younger cohort are coded as zero, which stands for either a non-MSA or an unidentified locality. Generally, these localities are rural areas. In the full sample, I classify these observations into the “small city” category.
purpose of constructing the wage growth measure in the regression corresponding to specification (5), I estimate the city-specific wage rate, \( \log w_k \), using the full sample and excluding \( T_i \times P_s \) and \( P_s \) in specification (4).

### 1.4.2 The learning effect

The results corresponding to specification (5) are displayed in Table 1.4. Column 1 reports the regression results using the full sample of age group 27-31. The coefficients of \( Z_{1j} \), \( Z_{2j} \), and \( Z_{3j} \) are all significantly positive, which implies that city population size, human capital stock, and population density are positively associated with the average learning effect \( g_j \). In the regression, I standardize the MSA-level variables to have mean zero and a standard deviation of one. A one-standard-deviation increase in the logarithm of MSA population size correlates with a 2.3 percent increase in \( g_j \) (approximately 0.7 standard deviations); a one-standard-deviation increase in the share of college graduates in the labor force is associated with a 1.6 percent increase in \( g_j \) (approximately a half of a standard deviation); a one-standard-deviation increase in the population density is correlated with a 0.8 percent increase in \( g_j \) (approximately 0.24 standard deviations). The literature suggests that learning in cities occurs through social interactions with workers from the same industry or different industries (Lucas, 1988; Glaeser, 1999; Peri, 2001; Charlot and Duranton, 2004; Moretti, 2004; Davis and Dingel, 2012). These kinds of interactions are likely to be more intense in bigger cities, probably due to higher population densities and larger diversities. Moreover, the high-end business service sector is usually concentrated in bigger cities, in which personal interactions are considered to
be more frequent than in the other industries (Arzaghi and Henderson, 2008). Furthermore, cities with a higher share of highly-educated workers in the labor force presumably have greater human capital spillovers, thereby enhancing the flow of ideas (Glaeser, et al. 1992). By interacting with peers doing similar work in other companies, young workers can acquire skills and end up being more productive; by socializing with a wide variety of people, young workers can attain knowledge about their own comparative advantages, thereby matching themselves with the best employers, the best careers, or the best occupations. Besides knowledge exchanges in social interactions, Brueckner (2011) discusses the other possible channels. For example, big cities provide a large labor pool in a particular industry, which gives employers a broad range of choice in hiring decisions, thereby making it harder for any individual to secure an entry-level job in this industry. Young workers therefore have an incentive to improve their skills through additional training. Furthermore, when socializing with employees working for other firms in their industry, workers may then judge their achievements against those of friends in other firms. This comparison may spur harder work as employees try to “look good” in the eyes of a broader social set.

For the individual whose MSA of current residence (city \(k\)) is coded as zero, the estimate of the entry wage rate, \(\log \hat{w}_k\), is an average wage rate for all non-MSA areas and unidentified localities in the state of current residence. This geographic unit could cover a broad region, in which multiple local labor markets may exist. Therefore, the wage growth measure is less accurate for this group of individuals. Column 2 of Panel A
reports the results of a subsample that drops individuals whose MSAs of current residence are coded as zero. The regression results remain stable.

Based on the estimates of $\gamma_1$, $\gamma_2$, and $\gamma_3$ in column 1, I calculate the average learning effects of 281 MSAs in my sample. In addition, 46 state-specific rural dummies capture the average learning effects of non-MSAs within states in which people entered the labor force. On average, college graduates starting work in big cities experience 7.2 percent higher wage growth over the five-year period immediately following labor force entry than their counterparts who started in small cities or rural areas, which is about two standard deviations of the estimated city’s average learning effect.

Panel B of Table 1.4 reports the estimates for the rates of dynamic return to learning ability in different-sized cities ($\pi_i$ and $\pi_o$). Applying the Delta Method, I calibrate $\pi_i\sigma_a$ and $\pi_o\sigma_a$. The point estimates for $\pi_i\sigma_a$ in columns 1 and 2 of Panel B are 0.042 and 0.047, respectively, at a one percent significance level, while the point estimates of $\pi_o\sigma_a$ are insignificant with a relatively small magnitude. College graduates exhibit substantial wage growth dispersion in big cities: a one-standard-deviation increase in learning ability causes nearly a 5 percent wage growth increase, given normality. However, such dynamic return to learning ability is absent in small cities and rural areas. When reversing the ranking of $\pi_j$, the regression results would imply a negative rate of dynamic return to learning ability in big cities, which contrasts with the observations that individuals endowed with higher learning abilities are more able to acquire skills and knowledge.

---

19 I place $\delta_0$ at 1.582, $\delta_1$ at -1.910, and $\mu$ at zero when calibrating $\pi_i\sigma_a$ and $\pi_o\sigma_a$. 

26
The parameter estimates support my prior belief that learning ability is better rewarded in big cities than in small cities and rural areas.

The individual whose MSA of residence five years ago (city \( j \)) is coded as zero lived either in a non-MSA or an unidentified locality in an identified state (rural areas). In columns 1 and 2, I classify them into the “small city” category. For these individuals, the MSA-level variables are unavailable, and state-specific rural dummies are included in columns 1 and 2 to capture the average learning effects in these localities. In column 3, I drop these individuals from the subsample used in column 2. The results are stable across columns 1 through 3.

Using the subsample of column 2, columns 4 and 5 present the regression results of movers (who changed MSAs of residence during the past five years) and stayers, respectively. The estimates of coefficients are not statistically different across these two subgroups, suggesting that the switching between periods is likely to be independent of learning ability, and positive sorting only occurs at labor force entry. (Column 6 in Panel A of Table 1.4 reports the t-statistic for testing the null hypothesis that the coefficients in columns 4 and 5 equal each other.) Furthermore, evidence from movers supports the statement that the related human capital accumulated in period one is portable when workers mover across cities.

The empirical results carry the following implications. First, young workers with higher learning ability experience greater wage growth in big cities (\( \pi_i > 0 \)), while such dynamic return to learning ability is absent in small cities and rural areas (\( \pi_0 = 0 \)). These
properties of the learning effects of cities in the 2000s echoes the study of Baum-Snow and Pavan (2012), which finds that the increases in the returns to unobserved skill in larger cities are more rapid than in small cities since the 1980s, thereby generating a more rapid growth in wage inequality in larger cities. Second, learning ability does not generate static heterogeneity in wages; that is, learning ability does not affect wages before recent college graduates start to acquire skills and knowledge in cities following labor force entry ($\mu = 0$).

Additionally, the geographic locations of birth states relative to available destinations for labor force entry affect wage growth by influencing the decision where one first enters the labor market. “Urban” states house or are directly adjacent to the smart cities such as New York City, Boston, Chicago and Washington, DC, while “rural” states are largely populated with rural areas or cities ranked at the bottom in terms of the city’s average learning effect. Given the actual destinations at which recent college graduates entered the labor force, for college graduates born in “urban” states like DC, NJ, NY, MD and MA, the average wage growth (over the five-year period immediately following labor force entry) about 6 to 7 percent higher than in states “rural” such as IA, SD, AR, WV and MT.

1.4.3 Evidence of positive sorting at labor force entry

Among previous studies, there is no consensus on the role of sorting based on unobserved ability in city size wage premium. Combes et al. (2008) claim that individual skills account for a large fraction of existing spatial wage disparities. The other studies suggest that the urban wage premium does not seem to be the result of omitted ability variables.
Based on the estimates for parameters that characterize the learning effect, I provide evidence of positive sorting into big cities at labor force entry. Sorting by unobserved learning ability contributes to the observed city-size wage premium by inflating the city-size wage growth difference.

In the U.S., positive sorting by learning ability into big cities when entering the labor force is an especially salient feature for college graduates born in "rural" states. To show this pattern, I categorize all birth states into three categories by the share of college graduates in big cities at labor force entry: below 0.30 (low-P state), between 0.30 and 0.47 (mid-P state), and above 0.47 (high-P state). None of the 15 states in the first category has a big city; four of the 16 in the second category, and all 18 in the last category have a big city. Figure 1.4 graphs the average adjusted wage growth \( \Delta \log w_{jk} - g_j \), for each city-size category at labor force entry, repeated for three categories of birth states. (I control for individual characteristics when calculating this adjusted wage growth. I standardize the adjusted wage growth so that individuals from high-P states in the “small city” category have zero average adjusted wage growth.) After controlling for the city’s average learning effect, a significantly positive city-size wage growth gap for states in the first two categories highlights the role of positive sorting into big cities by unobserved heterogeneity among individuals born in “rural” states.

---

20 States in the first category are SC, NC, KY, MS, TN, AR, AL, WV, UT, LA, OK, NE, WI, ID, and MT. The second category includes IA, NH, NV, SD, RI, ND, DE, KS, ME, CT, VT, WY, OH, NM, IN, VA, and GA. The third category covers MI, NJ, TX, PA, FL, WA, MN, MO, NY, CO, OR, MA, IL, AZ, CA, MD, and DC.

21 Using the NLSY79, Baum-Snow and Pavan (2012) find slight negative sorting into both medium and large cities at labor force entry for college graduates. The sample of the NLSY79 is designated to be representative of youths living in the United States in 1979, yet within a city at labor force entry, it is not
Moreover, the monotonic relationship between the average wage growth in the “big city” category and the share of college graduates in big cities at labor force entry features the source for identifying the selection at labor force entry.

Since $\mu = 0$, $\pi_0 = 0$, and $\pi_1 > 0$, the adjusted wage growth, $(\Delta \log w_{jk} - g_j)$, equals $\pi_1 \times a$ plus a stochastic term for individuals entering the labor force in big cities, while for individuals in small cities and rural areas at labor force entry, this measure equals only a stochastic term. Therefore, in small cities and rural areas the full distribution of $(\Delta \log w_{jk} - g_j)$ should overlap across birth states; in big cities the full distribution of $(\Delta \log w_{jk} - g_j)$ would have a fat right tail resulting from positive sorting by learning ability, and the distribution in a state of a larger selection threshold would be more heavily tailed. As an example, the top two panels in Figure 1.5 graph the full distribution of $(\Delta \log w_{jk} - g_j)$ of individuals from two geographically adjacent birth states, Alabama and Georgia. As a typical “rural” state, Alabama is expected to have a larger selection threshold than Georgia, in which a big city, Atlanta, is located. The bottom two panels in Figure 1.5 graph the simulated counterparts, using selection thresholds inferred from the observed $P_s$ of these two states. The patterns from the actual data approximate the simulated ones as expected.

In a two-city, two-period world, the observed city-size wage premium in period two is necessarily a random sample across birth states given the limited sample size. In big cities, individuals from states far away from a big city tend to be under-sampled and thereby less representative in terms of learning ability. Therefore, any sorting pattern may potentially emerge using a sample from that data set.
\[(\log w_i - \log w_o) + (g_i - g_o) + (\pi_1E(a \mid a > a^\star) - \pi_0E(a \mid a \leq a^\star))\]

assuming that individuals do not switch cities between periods. Besides the city-size wage level premium \((\log w_i - \log w_o)\), the city-size wage growth gap \((g_i - g_o)\) and sorting by learning ability at labor force entry \((\pi_1E(a \mid a > a^\star) - \pi_0E(a \mid a \leq a^\star))\) also influence the observed city-size wage premium. Comparing big and small cities in the sample of 281 MSAs, the city-size wage level premium is about 15.2 percent; the city-size wage growth gap is around 7.2 percent; sorting by learning ability generates another 3.1 percent wage difference, assuming normality. Together, they result in a 25.5 percent observed city-size wage premium.

1.4.4 Difference in the distribution of learning ability among birth states

The baseline specifications are based on the assumption that recent college graduates from different birth states share a single learning ability distribution. That is, the main identification assumption is that the error terms in equations (4) and (5) do not capture any unobserved birth-state characteristics that correlate with both wages and the share of college graduates in big cities at labor force entry. In this section, I explicitly incorporate the possible difference in the distribution of learning ability among birth states into the estimation specifications and show that the general implications remain robust.

Several factors may cause a difference in the distribution of learning ability among birth states. One major concern is that the unobserved heterogeneity in parents may differ across birth states. For example, parents with higher unobserved ability may sort into “urban” states. If intergenerational transfer in unobserved ability exists, college graduates
born in “urban” states on average would have higher learning abilities than college graduates born in “rural” states. Differences in infrastructure and school system may also contribute to the differentials in average learning ability across birth states. In fact, state-level average test scores from the National Assessment of Education Progress (the NAEP) show that “urban” states, such as MA, MD, NJ, and TX, have higher average test scores than “rural” states, such as WV, MS, AL, AR, TN, SC, and NC. Additionally, parents in “urban” states have greater education attainment, measured by the share of college degree holders among ever married men aged between 40 and 60 in 2000, than in “rural” states. To incorporate this difference among birth states, I introduce a birth-state-specific mean of learning ability, denoted as $m_S$, where $S$ indexes birth state, into the model. The observed variation in the share of college graduates in big cities at labor force entry ($P_S$) could result from either a difference in the selection threshold ($a_S^*$) or a shift in the mean of learning ability ($m_S$) among birth states. For instance, “urban” states may have smaller selection thresholds and larger means of learning ability, thereby leading to larger shares of college graduates in big cities at labor force entry. If we ignore the difference in $m_S$, in each city-size category the average learning ability of people born in these “urban” states would be underestimated.

To show how the difference in the mean of learning ability among birth states affects my estimation results, I use a linear function to approximate the relationship between $m_S$ and

---

22 I assume that test scores are partially determined by learning abilities. I document the average reading (49 states in 2011), math (49 states in 2011), writing (44 states in 2007), and science (45 states in 2009) scores for white students in grade 8 from the NAEP, which provides state-level average scores for public school students within a single assessment year.
when calculating the selected means: \( m_s \approx \kappa_0 + \kappa_1 \times P_s \).

Here, \( \kappa_1 \) is positive. (The share of college graduates entering the labor force in big cities is positively correlated with state-level average test scores as well as the share of college degree holders among ever married men aged between 40 and 60) Panel B in Table 1.2 presents the new relationships between the key coefficients in the regression specifications and parameters from the model, incorporating the birth-state-specific mean of learning ability. The zero estimate of \( \mu \) is robust regardless of \( m_s \); the zero estimate of \( \pi_0 \) would still be robust unless \((\kappa_1 - \sigma_0 \delta)\) is zero; the estimate of \( \pi, \sigma_a \) equals \( \pi, \sigma_a + (\kappa_1 / \delta_1) \pi_1 \), which is smaller than \( \pi, \sigma_a \), since \( \delta_1 < 0 \), and \( \kappa_1 > 0 \). Therefore, the possible difference is likely to cause me to underestimate the rate of dynamic return to learning ability in big cities. In addition, I directly add \( m_s \) and its interaction with \( T_i \) into specification (5). I use the state’s average reading score as well as the share of college degree holders among ever married men aged between 40 and 60 as proxies for \( m_s \), separately. Table 1.5 presents the results. While the estimates of \( \pi_0, \sigma_a \) remain zero across specifications, the estimate of \( \pi, \sigma_a \) increases when additional birth-state controls are included as expected.

1.4.5 City-size cutoff

In the main analysis, I use 1.5 million as the cutoff for defining big cities. An appropriate city-size cutoff matters for estimation. The top two panels in Figure 1.6 present the 95 percent confidence intervals of estimated \( \pi, \sigma_a \) and \( \pi_0, \sigma_a \), using different city-size cutoffs.

\begin{align*}
E(a \mid a > a^*_S) &= m_s + \sigma_a \tilde{\lambda}_t (\Phi^{-1}(1 - P_s)) \approx (\kappa_0 + \kappa_1 \times P_s) + \sigma_a (\tilde{\lambda}_t + \delta_1 \times P_s) , \\
E(a \mid a < a^*_S) &= m_s + \sigma_a \tilde{\lambda}_0 (\Phi^{-1}(1 - P_s)) \approx (\kappa_0 + \kappa_1 \times P_s) - \sigma_a (\tilde{\lambda}_0 + \delta_0 \times P_s) .
\end{align*}

\[ 23 \]
from 0.5 to 2.1 million. For the city-size cutoffs valued above 1.2 million, the point estimate of $\pi_1\sigma_a$ remains positive at a five percent significance level (top left panel in Figure 1.6). The confidence intervals of $\pi_0\sigma_a$ all cover zeros, and the standard error generally decreases with the value of the city-size cutoff through 1.5 million (top right panel in Figure 1.6). The third panel in Figure 1.6 graphs the t-ratios of the difference between $\pi_1\sigma_a$ and $\pi_0\sigma_a$ against different city-size cutoffs. At the 1.5-million this value peaks at 1.81. The patterns of parameter estimates in Figure 1.6 suggest that city-size cutoffs below 1.2 million may put cities that are commonly deemed as small ones into the “big city” category, leading to an underestimation of $(\pi_1\sigma_a)$. The standard error of $\pi_1\sigma_a$ generally increases when dropping a city at the bottom of the city-size ranking each time from the “big city” category, since a larger cutoff leaves fewer birth states with smaller sizes of observations in the corresponding “big city” category. Therefore, it is reasonable to use the 1.5-million cutoff for classifying cities into two city categories.

1.4.6 Role of natural endowments

In the baseline specification (Equation 5), a city’s average learning effect is correlated with only the city’s human endowments (i.e., population size, population density, and human capital stock). Albouy (2010) suggests that natural endowments, such as sunshine and coastal proximity, may have substantial effects on firm productivity. These related natural amenities could also affect the extent to which a city promotes wage growth. For example, extremely hot summers or cold winters may appear to be bad for social activities, thereby inhibiting wage growth. In this section, I re-run the regression corresponding to Equation 5 by further including the MSA-level natural amenity
variables (i.e., precipitation, coastal proximity, heating-degree days, cooling-degree days, and January temperature). Columns 1 and 2 in Table 1.5 present the estimated coefficients of the MSA-level variables, using the subsample corresponding to column 3 in Table 1.4. For the purpose of comparison, I also regress the city-specific wage rate (entry wage level) estimated in Section 1.4.1 on the sets of human and natural amenity variables used in columns 1 and 2 of Table 5. I report the corresponding results in columns 3 and 4 of Table 1.5. The results suggest that a city’s human endowments enhance both wage level and wage growth, while the natural endowments, such as favorable location, seem to affect only the wage-level differences across cities.

1.5 Evidence from the NLSY79

This section shows that panel data on young adults reveal an experience profile of wages that is consistent with my results reported above. I use the sample drawn from a confidential version of the 1979 National Longitudinal Survey of Youth (the NLSY79) covering the period from 1979 through 1994 that provides information on metropolitan statistical area of residence for each respondent in each round of the survey as well as the AFQT score as a proxy for learning ability. Given the properties of learning effects, I expect the following patterns from the NLSY79. First, for college graduates starting in big cities, their earnings-experience profiles are, on average, steeper than individuals who enter the labor force in small cities. Second, at the time of first entering, differential

---

24 Climate variables are from the County and City Data Book 2000 provided by the U.S. Census Bureau. Coastal proximity is measured by the distance to the nearest water (Ocean and Great Lakes) calculated by ArcGIS software.

25 The NLSY79 is a nationally representative sample of men and women who were between the ages of 14 and 22 when they were selected at the beginning of 1979. The AFQT, or the Armed Force Qualification Test, consists of the following four sections from the Armed Services Vocational Aptitudes Battery: word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge.
AFQT scores do not differentially affect wage rates. Third, for individuals entering the labor force in big cities, higher AFQT scores associate with steeper earnings-experience profiles, while for individuals entering the labor force in small cities or rural areas, the slope of the earnings-experience profile does not vary by the AFQT score.

I construct information on wages and migration histories for a sample of white men from the time of their entry into the labor force. I include only individuals who reported a bachelor’s degree as the highest degree he had ever received by 1994. The sample for estimation has 1,974 observations of 258 people. Nearly 90 percent of individuals in the sample have over six years of work experience. MSAs with a population above 1.2 million in the 1990 census are defined as big cities. In my sample, 144 individuals entered the labor force in a big city. AFQT scores are standardized to have mean zero and a standard deviation of one. Real wages in 1984 prices were calculated using CPI deflators from the U.S. Census Bureau.

To show the wage patterns from the raw data, I divide the sample into two subgroups of individuals by their AFQT scores: individuals with AFQT scores above the median are defined as high-type; others are defined as low-type. For these two subgroups of individuals and for two city-size categories at labor force entry, Figure 1.7 plots hourly wages.

---

26 The sample includes 494 white men who received their bachelor’s degrees prior to 1994 from the NLSY79 random sample of 3,003 men. 46 individuals are dropped because they were in the military at some point. An additional 10 individuals are dropped because they received their bachelor’s degrees prior to 1979. Only jobs after an individual has received a bachelor’s degree while not enrolling in any school are kept in the sample. The first year that an individual appears in the sample is when he entered the labor force. I keep individuals who entered the labor force prior to age 30. 90 percent of them entered the labor market between the age of 22 and 26. I drop individuals who miss wage rate information by the time of entering the labor force. Individuals with missing AFQT scores are dropped. Observations whose real wages were below $1 or above $100 are dropped. Finally, I drop individuals whose locations at labor force entry are unidentifiable. The resulting sample has 1,974 observations of 258 individuals.

27 Individuals with either the first or the second observation appearing in an MSA with a population above 1.2 million are treated as entering the labor force in big cities.
wage rates against years of experience in the labor force. First, as expected, entering the labor force in big cities is associated with a significant wage growth premium. Second, the earnings profiles of two subgroups seem to diverge from each other only for individuals who enter the labor force in big cities (right panel of Figure 1.7), which is consistent with the property of the learning effect: learning ability is positively rewarded only in big cities. I estimate the following specification to see whether the patterns from the raw data are robust with full controls.

$$\log w_{it} = \beta_0 \exp_{it} + \beta_1 \exp_{it} \times T_i + \pi_0 (AFQT_i \times \exp_{it}) + \pi_1 (AFQT_i \times \exp_{it} \times T_i) + \mu(AFQT_i \times enterLF_{it}) + \delta' \Omega + \varepsilon_{it}$$

where $\log w_{it}$ is the log of real hourly wage of individual $i$ in year $t$, $T_i$ indicates whether individual $i$ enters the labor force in a big city, $\exp_{it}$ is the experience level in year $t$, $enterLF_{it}$ indicates whether individual $i$ enters the labor force in year $t$, and $\varepsilon_{it}$ is the error term. The vector $\Omega$ includes an indicator for urban residence in year $t$, the individual’s AFQT score, year fixed effects, and a cubic polynomial in experience. Table 1.7 presents the regression results. Column 1 includes only the experience level and its interaction term with the city-size dummy at labor force entry. The coefficient of the interaction between experience and the city-size dummy is significantly positive in both columns 1 (0.15 with $t$ at 4.19) and 2 (0.12 with $t$ at 3.54). This suggests the experience profile of wages is steeper for college graduates who started work in big cities than in small cities. In column 2, while the coefficient of the interaction between AFQT and experience is insignificant with a small magnitude (-0.004 with $t$ at -1.01), the coefficient of the interaction between AFQT, experience, and the city-size dummy at labor force
entry is positive and statistically significant (0.013 with t at 3.63). The results indicate that unobserved ability generates dispersion in wage growth only for college graduates who started work in big cities. Finally, the AFQT coefficient is not significantly different from zero at the time of first entering the labor force ($AFQT_i \times enterLF_t$), suggesting that before college graduates acquire skills and knowledge, unobserved ability does not affect wages. Compared with column 1, including additional ability controls in column 2 significantly lowers the coefficient of the interaction between experience and the city-size dummy, suggesting that positive sorting by AFQT scores into big cities occurred at labor force entry. Columns 3 and 4 in Table 1.7 present the regression results of movers (who changed MSAs of residence at least once after labor force entry) and stayers. Estimates are statistically the same between these two subgroups, which is consistent with the evidence from the census and suggests that the wage growth gain is portable for workers who moved across cities. (Column 5 of Table 1.7 reports the t-statistic for testing the null hypothesis that the coefficients in columns 4 and 5 equal each other.) Overall, the results from the NLSY79 reveal an experience profile of wages that is consistent with my results reported in Section 1.4.

1.6 Conclusions

This paper studies the city’s wage growth attribute—the learning effect. Today, especially in the developed world, the most important knowledge of skilled people is usually acquired after graduation (Glaeser, 2011). This paper suggests that locations where college graduates first enter the labor force play an important role in facilitating the acquisition of skills and knowledge and promoting subsequent wage growth.
Furthermore, an individual’s learning ability, which is presumably associated with social skills as well as cognitive aptitudes, begets further learning in big cities. Because of that a self-selection by learning ability would occur at labor force entry. This paper provides empirical evidence that positive sorting by learning ability into big cities when first entering the labor force is a salient feature for college graduates born in “rural” states in the United States. Estimates imply that in addition to the city-size wage level premium of 15.2 percent, college graduates starting work in big cities experienced 7.2 percent higher wage growth over the five-year period immediately following labor force entry than their counterparts in small cities. Positive sorting by learning ability adds another 3.1 percent to the wage growth difference between big and small cities.

Several mechanisms could potentially contribute to the more rapid learning in larger cities. First, social interactions are likely to be more intense in bigger, denser, and more highly educated cities. By interacting with peers doing similar work in other companies, young workers can acquire skills and end up being more productive; by socializing with a wide variety of people, young workers can attain knowledge about their own comparative advantages, thereby matching themselves with the best employers, the best careers, or the best occupations. These kinds of interactions tend to require social/people skills from young workers, which is consistent with the empirical findings of Bacolod et al. (2009): cities raise the prices of cognitive and people skills, not the prices of physical skills such as strength and motor skills. Future work will attempt to understand more about the mechanisms through which people learn in cities.
## 1.7 Tables

### Table 1.1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>27-31</th>
<th>22-26</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Migration history</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I(start from big city)</td>
<td>0.426</td>
<td>0.490</td>
</tr>
<tr>
<td>I(switch)</td>
<td>0.437</td>
<td></td>
</tr>
<tr>
<td><strong>Labor market information</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>weeks worked in 1999</td>
<td>51.257</td>
<td>50.865</td>
</tr>
<tr>
<td>(2.207) (2.742)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>usual hours worked per week in 1999</td>
<td>46.873</td>
<td>45.452</td>
</tr>
<tr>
<td>(8.383) (8.027)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hourly wage rate</td>
<td>21.871</td>
<td>16.316</td>
</tr>
<tr>
<td>(15.745) (9.867)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Industry sectors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I(construction)</td>
<td>0.041</td>
<td>0.041</td>
</tr>
<tr>
<td>I(manufacturing)</td>
<td>0.164</td>
<td>0.145</td>
</tr>
<tr>
<td>I(transportation and warehousing)</td>
<td>0.026</td>
<td>0.020</td>
</tr>
<tr>
<td>I(utilities)</td>
<td>0.028</td>
<td>0.024</td>
</tr>
<tr>
<td>I(wholesale)</td>
<td>0.054</td>
<td>0.040</td>
</tr>
<tr>
<td>I(retail trade)</td>
<td>0.113</td>
<td>0.110</td>
</tr>
<tr>
<td>I(business services)</td>
<td>0.275</td>
<td>0.284</td>
</tr>
<tr>
<td>I(personal services)</td>
<td>0.019</td>
<td>0.021</td>
</tr>
<tr>
<td>I(entertainment and recreation services)</td>
<td>0.032</td>
<td>0.035</td>
</tr>
<tr>
<td>I(professional and related)</td>
<td>0.168</td>
<td>0.202</td>
</tr>
<tr>
<td>I(public administration)</td>
<td>0.079</td>
<td>0.078</td>
</tr>
<tr>
<td><strong>Other demographics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>28.882</td>
<td>24.377</td>
</tr>
<tr>
<td>(1.381) (1.301)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>single</td>
<td>0.388</td>
<td>0.668</td>
</tr>
<tr>
<td>observations</td>
<td>42958</td>
<td>35063</td>
</tr>
</tbody>
</table>

Notes: Standard deviations are in parentheses.
Table 1.2: Regression Coefficients and Structural Parameters

Panel A

<table>
<thead>
<tr>
<th>Equation</th>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$T_i \times P_S$</td>
<td>$\varphi_1 = (\mu \sigma_a \delta_1 + \mu \sigma_a \delta_0)$</td>
</tr>
<tr>
<td>4</td>
<td>$P_S$</td>
<td>$\varphi_2 = -\mu \sigma_a \delta_0$</td>
</tr>
<tr>
<td>5</td>
<td>$T_i \times P_S$</td>
<td>$\alpha_2 = (\mu + \pi_1) \sigma_a \delta_1 + (\mu + \pi_0) \sigma_a \delta_0$</td>
</tr>
<tr>
<td>5</td>
<td>$P_S$</td>
<td>$\alpha_3 = -(\mu + \pi_0) \sigma_a \delta_0$</td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th>Equation</th>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$T_i \times P_S$</td>
<td>$\varphi_1 = (\mu \sigma_a \delta_1 + \mu \sigma_a \delta_0)$</td>
</tr>
<tr>
<td>4</td>
<td>$P_S$</td>
<td>$\varphi_2 = \mu (\kappa_1 - \sigma_a \delta_0)$</td>
</tr>
<tr>
<td>5</td>
<td>$T_i \times P_S$</td>
<td>$\alpha_2 = (\mu + \pi_1) (\kappa_1 + \sigma_a \delta_1) - (\mu + \pi_0) (\kappa_1 - \sigma_a \delta_0)$</td>
</tr>
<tr>
<td>5</td>
<td>$P_S$</td>
<td>$\alpha_3 = (\mu + \pi_0) (\kappa_1 - \sigma_a \delta_0)$</td>
</tr>
</tbody>
</table>
Table 1.3: Static Heterogeneity of Learning Ability

Panel A. log(hourly wage rate), 22-26 cohort

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_i \times P_S$</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>$P_S$</td>
<td>-0.010</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>observations</td>
<td>35063</td>
<td>28798</td>
</tr>
</tbody>
</table>

Panel B. Parameter estimates from the structural model

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu\sigma_{\delta_1}$</td>
<td>-0.008</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>$\mu\sigma_{\delta_0}$</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.027)</td>
</tr>
</tbody>
</table>

Notes:*significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses, and they are clustered by MSA of current residence and birth state. $P_S$ is the percentage of college graduates born in state $S$ and entered the labor force in big cities (calculated from age group 22-26). $T_i$ indicates whether the individual entered the labor force in a big city. Column 1 includes 339 MSA dummies (292 MSAs and 47 rural areas). Column 2 excludes individuals entering the labor force in a non-MSA or an unidentified locality. They also include age, age squared, marital status, and industrial dummies.
Table 1.4: Main Results and Parameter Estimates from the Structure Model

<table>
<thead>
<tr>
<th>Panel A. Wage growth $\Delta \log w_{ijk}$, 27-31</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>ln(population size)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>share college graduates in labor force</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>population density</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$T_i$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$T_i \times P_S$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$P_S$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>R2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>observations</td>
</tr>
</tbody>
</table>

**Panel B. Parameter estimates from the structural model**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1 \sigma_a$</td>
<td>0.042***</td>
<td>0.047***</td>
<td>0.047***</td>
<td>0.048</td>
<td>0.045**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.030)</td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>$\pi_0 \sigma_a$</td>
<td>0.015</td>
<td>0.007</td>
<td>0.004</td>
<td>0.000</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.017)</td>
<td>(0.020)</td>
<td>(0.023)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses, and they are clustered by MSA of residence 5 years ago and birth state. $P_S$ is the percentage of college graduates born in state $S$ and entered the labor force in big cities (calculated from age group 22-26). subsample1 excludes individuals whose MSAs of current residence are coded as zero. subsample2 excludes individuals whose MSAs of current residence or residence 5 years ago are coded as zero. movers indicates the individual whose MSA of current residence differs from the MSA of residence 5 years ago; otherwise, individuals are stayers. I place $\mu$ at zero when calibrating $\pi_1 \sigma_a$ and $\pi_0 \sigma_a$. The regression in column 1 includes a dummy indicating the MSA of current residence is coded as zero. For individuals whose MSAs of residence 5 years ago are coded as zero, columns 1 and 2 include 46 state-specific rural dummies to capture the average learning effect in rural areas within a state. All columns include age, age squared, marital status, industrial dummies, and a constant term.
Table 1.5: Additional Birth-State Controls

**Panel A. Wage growth** $\Delta \log w_{ijk}, 27-31$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>full sample</td>
<td>full sample</td>
<td>full sample</td>
<td>subsample1</td>
<td>subsample1</td>
<td>subsample1</td>
</tr>
<tr>
<td>ln(population size)</td>
<td>0.023***</td>
<td>0.024***</td>
<td>0.024***</td>
<td>0.024***</td>
<td>0.024***</td>
<td>0.024***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>share college graduates in labor force</td>
<td>0.016***</td>
<td>0.014***</td>
<td>0.014***</td>
<td>0.015***</td>
<td>0.015***</td>
<td>0.015***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>population density</td>
<td>0.008***</td>
<td>0.007***</td>
<td>0.007***</td>
<td>0.008***</td>
<td>0.008***</td>
<td>0.008***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>$T_i \times P_s$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>$P_s$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>$T_i \times \frac{40\text{to}60\text{collegeedu}}$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>$\frac{40\text{to}60\text{collegeedu}}$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>$T_i \times \text{NAEPreading}$</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>NAEPreading</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>observations</td>
<td>42958</td>
<td>42958</td>
<td>42958</td>
<td>35262</td>
<td>35262</td>
<td>35262</td>
</tr>
</tbody>
</table>

**Panel B. Parameter estimates from the structural model**

<table>
<thead>
<tr>
<th></th>
<th>$\pi_0 \sigma_a$</th>
<th>$\pi_0 \sigma_a$</th>
<th>$\pi_1 \sigma_a$</th>
<th>$\pi_1 \sigma_a$</th>
<th>$\pi_1 \sigma_a$</th>
<th>$\pi_1 \sigma_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.015</td>
<td>0.012</td>
<td>0.010</td>
<td>0.016</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>$\pi_0 \sigma_a$</td>
<td>0.015</td>
<td>0.012</td>
<td>0.018</td>
<td>0.007</td>
<td>0.002</td>
<td>0.012</td>
</tr>
<tr>
<td>$\pi_0 \sigma_a$</td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses, and they are clustered by MSA of residence 5 years ago and birth state. subsample1 excludes individuals whose MSAs of current residence are coded as zero. The regression corresponding to columns 1, 2, and 3 include a dummy indicating the MSA of current residence is coded as zero. For individuals whose MSAs of residence 5 years ago are coded as zero, all columns include state-specific rural dummies to capture the average learning effect in rural areas within a state. All columns include age, age squared, marital status, industrial dummies, and a constant. frac40to60collegeedu is the share of college graduates among married males aged between 40 and 60 in 2000 in each state. NAEPreading is the state’s reading average reading test scores for white students in grade 8 in public schools in 2011.
Table 1.6: Roles of Human and Natural Endowments

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>$g_{i1}$</th>
<th>$g_{i2}$</th>
<th>$\log w_{i3}$</th>
<th>$\log w_{i4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(population size)</td>
<td>0.024***</td>
<td>0.023***</td>
<td>0.044***</td>
<td>0.043***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>share college graduates in labor force</td>
<td>0.015***</td>
<td>0.014***</td>
<td>0.037***</td>
<td>0.038***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>population density</td>
<td>0.008***</td>
<td>0.007***</td>
<td>0.020***</td>
<td>0.021***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Distance to closest water</td>
<td>-0.564</td>
<td>-0.932</td>
<td>-11.222***</td>
<td>-12.282***</td>
</tr>
<tr>
<td>(Ocean or Great Lakes)</td>
<td>(2.931)</td>
<td>(2.912)</td>
<td>(3.726)</td>
<td>(4.081)</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>0.321</td>
<td>0.289</td>
<td>-0.540</td>
<td>-0.685</td>
</tr>
<tr>
<td>(1,000,000 inches)</td>
<td>(0.302)</td>
<td>(0.303)</td>
<td>(0.574)</td>
<td>(0.576)</td>
</tr>
<tr>
<td>Heating-Degree Days</td>
<td>0.000</td>
<td></td>
<td>-0.006</td>
<td></td>
</tr>
<tr>
<td>(1,000 days)</td>
<td>(0.002)</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooling-Degree Days</td>
<td>0.001</td>
<td></td>
<td>-0.009</td>
<td></td>
</tr>
<tr>
<td>(1,000 days)</td>
<td>(0.006)</td>
<td>(0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>January temperature</td>
<td></td>
<td>1.317</td>
<td>1.824</td>
<td></td>
</tr>
<tr>
<td>(1,000 degrees Fahrenheit)</td>
<td></td>
<td>(1.477)</td>
<td>(2.445)</td>
<td></td>
</tr>
<tr>
<td>January temperature squared</td>
<td></td>
<td>-17.619</td>
<td>-17.334</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20.266)</td>
<td>(31.219)</td>
<td></td>
</tr>
</tbody>
</table>

Notes:*significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. $g_{i1}$ represents the city's average learning effect. $\log w_{i3}$ represents the city-specific wage rate. Climate variables are from normal values for the 30-year period, 1961-1990. January temperature is the average daily temperature in January. One heating degree day is accumulated for each whole degree that the mean daily temperature is below 65 degree Fahrenheit. One cooling degree day is accumulated for each whole degree that the mean daily temperature is above 65 degrees Fahrenheit. Distance to closest water is the relative distance calculated by ArcGIS. ln(population size), share college graduates in labor force, and population density are standardized to have mean zero and a standard deviation of one. Regressions in columns 3 and 4 are weighted by the number of observations in each MSA in the entry wage regression of the younger cohort.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>t-stat (3=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>full</td>
<td>full</td>
<td>movers</td>
<td>stayers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\exp_{it}$</td>
<td>0.134***</td>
<td>0.133***</td>
<td>0.126***</td>
<td>0.147***</td>
<td>-0.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.031)</td>
<td>(0.037)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\exp_{it} \times T_i$</td>
<td>0.015***</td>
<td>0.012***</td>
<td>0.012***</td>
<td>0.011**</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AFQT_i \times \exp_{it}$</td>
<td>-0.004</td>
<td>-0.005</td>
<td>0.000</td>
<td></td>
<td>-0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AFQT_i \times \exp_{it} \times T_i$</td>
<td>0.013***</td>
<td>0.014***</td>
<td>0.010*</td>
<td></td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AFQT_i \times enterLF_{it}$</td>
<td>0.007</td>
<td>-0.015</td>
<td>0.056</td>
<td></td>
<td>-0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.045)</td>
<td>(0.080)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.165</td>
<td>0.175</td>
<td>0.151</td>
<td>0.272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>1974</td>
<td>1974</td>
<td>1,417</td>
<td>557</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. $\exp_{it}$ represents the experience in year $t$. $T_i$ indicates whether individual $i$ enters the labor force in big cities. $enterLF_{it}$ indicates the time of first entering the labor force. All columns include the individual's AFQT score, an indicator for urban residence in year $t$, year fixed effects, and a cubic polynomial in experience. Column 3 uses the subsample of movers who changed MSAs of residence at least once after labor force entry. Column 4 uses the subsample of stayers who never changed MSAs of residence after labor force entry.
1.8 Figures

**Figure 1.1: Patterns in the Raw Data**

Notes: Panel A shows the average wage growth for each city-size category at labor force entry. I standardize the wage growth measure so that individuals in the “small city” category have zero average wage growth. Panel B plots the average wage growth of individuals who were born in birth states with (urban states) and without (rural states) a big city. Panel C shows the average wage growth for each city-size category at labor force entry, repeated for urban states and rural states. In Panel B and Panel C, I standardize wage growth so that individuals from urban states in the “small city” category have zero average wage growth. Panel D draws the average entry hourly wage, calculated from the younger cohort (22-26), for people who were born in rural states and urban states, respectively. Panel E shows the average entry hourly wage, calculated from the younger cohort (22-26), for each city-size category at labor force entry, repeated for urban states and rural states. In Panel D and Panel E, I standardize the entry wage rate so that individuals from rural states in the “small city” category have zero average entry wage rate. small indicates that the individual enters the labor force in either a small MSA (less than 1.5 million people in 2000) or a non-MSA, and big represents that the individual enters the labor force in a big MSA (more than 1.5 million people in 2000).
Figure 1.2: Share in Big Cities at Labor Force Entry, by Birth State
Figure 1.3: Geographic Locations of Birth States and Destinations in the “Big City” Category

Notes: Blue circles represents MSAs of more than 1.5 million people in 2000. States in dark green are “rural” states. States in light green are “urban” states.
Figure 1.4: Positive Sorting at Labor Force Entry—Average of Adjusted Wage Growth

Notes: The graph shows the average wage growth, controlling for the city’s average learning effect and individual characteristics, for each city-size category at labor force entry, repeated for birth states in each interval classified based on the share of college graduates in big cities at labor force entry. I standardize the adjusted wage growth so that individuals from high-P states in the “small city” category have zero average adjusted wage growth. Small indicates that the individual enters the labor force in a small MSA (less than 1.5 million people in 2000) or a non-MSA, and big represents that the individual enters the labor force in a big MSA (more than 1.5 million people in 2000).
Figure 1.5: Positive Sorting at Labor Force Entry—Full Distribution of Adjusted Wage Growth

Notes: In the older cohort of the actual data, 80 of 559 individuals from AL went to big cities at LF entry, and 329 of 863 individuals from GA went to big cities at LF entry. In the simulated data, 68 of 559 individuals from AL enter the labor force in big cities, and 314 of 863 from GA enter the labor force in big cities. small indicates that the individual enters the labor force in a small MSA (less than 1.5 million people in 2000) or a non-MSA, and big represents that the individual enters the labor force in a big MSA (more than 1.5 million people in 2000).
Figure 1.6: Parameter Estimates using Different City-Size Cutoffs
Figure 1.7: Experience Profiles of Wages, the NLSY79
1.9 Appendix A: Selection Thresholds in the Baseline Equilibrium

For the individual endowed with learning ability \( a \), the expected lifetime utility of entering the labor force in city \( j \), denoted as \( E(V_j | a) \), is given by:

\[
E(V_j | a) = \Psi_j + \phi_a \mu a + \beta \sum_{k=0}^{J-1} \left\{ \Psi_k + \phi_1 (\mu a + g_j(a) + m_\varepsilon) \right\} \times P(\varepsilon_j + \Psi_k > \max \{ \varepsilon_{j-k} + \Psi_{-k} \})
\]

where \( m_\varepsilon \) is the mean of \( \varepsilon_{jk} \). Assuming that \( \varepsilon_{jk} \) are independently distributed (over \( j \) and over \( k \)) with the type I extreme value distribution, I rewrite \( E(V_j | a) \) as follows:

\[
E(V_j | a) = \Psi_j + \phi_a \mu a + \beta \sum_{k=0}^{J-1} \left\{ \Psi_k + \phi_1 (\mu a + g_j(a) + m_\varepsilon) \right\} \times \frac{\Psi_k}{\sum_{h=0}^{J-1} \Psi_h}
\]

\[
= \Psi_j + \beta \phi_1 g_j + \beta \phi_1 \pi_j a + \phi_1 (1 + \beta) \mu a + \beta \phi_1 m_\varepsilon + \frac{\sum_{k=0}^{J-1} \Psi_k}{\sum_{h=0}^{J-1} \Psi_h} \Psi_k^2
\]

Since \( \pi_{j-1} > \pi_{j-2} > \ldots > \pi_1 \geq \pi_0 \geq 0 \), \( \partial E(V_{j+1} | a) / \partial a > \partial E(V_j | a) / \partial a \geq 0 \), \( j = 0, \ldots, J-2 \). Therefore, a unique \( a_j^* \) exists, such that \( E(V_{j+1} | a_j^*) = E(V_j | a_j^*) \), \( j = 0, \ldots, J-2 \).

Furthermore, since every locality is populated in equilibrium, the segment of \( E(V_j | a) \) between \( a_{j-1}^* \) and \( a_j^* \) overlaps the upper envelope of the expected lifetime utilities in all localities.
1.10 Appendix B: Equilibrium Condition for Firms

Labor $L$ and capital $K$ enter as factors of production with input prices $w_j$ and $r$, respectively. In equilibrium the unit cost of producing the numeraire traded good at a locality is equal to its price as a result of the full mobility and the constant-returns-to-scale technology.

$$c(w_j, r; A_j) = \{\min Lw_j + Kr : F(K, L; A_j) = 1\} \equiv 1$$

Assuming that the rental rate of capital is the same in every locality, total differentiation derives the following equilibrium relationship between any two localities ($j$ and $k$), where $\varphi_L$ is firm’s expenditure share on labor.

$$\log w_j - \log w_k = \frac{\log A_j - \log A_k}{\varphi_L}.$$  

It implies that in equilibrium the differential in the local-productivity ($A$) determines the differential in the city-specific wage rate ($\log w$).
1.11 Appendix C: A Distance Model

In the real world, individuals share common sets of big and small cities: \( \{0, \ldots, J_0-1\} \) and \( \{J_0, \ldots, J-1\} \), where \( (J-J_0) \) represents the total number of big cities, and \( J_0 \) is the total number of small cities. Cities in each city-size category share the same \( \pi_j, \pi_1 \) and \( \pi_0 \) denote the rate of dynamic return to learning ability in a big city and a small city, respectively. I assume that \( \pi_1 > \pi_0 \).

Let \( d_{js} \) denote the distance between destination \( j \) and birth state \( S \). For the individual born in state \( S \) and working in destination \( j \), \( d_{js} \) enters the flow utility in the following way: \( \Psi_{js} = \phi_1 \log w_j + \phi_2 \log Q_j - \phi_3 \log p_j - \phi_4 \log d_{js} \), where \( \phi_4 > 0 \). Given birth state \( S \) and learning ability \( a \), the individual’s expected lifetime utility of entering the labor force in city \( j \), denoted as \( E(V_j | a, S) \), is given by:

\[
E(V_j | a, S) = \Psi_{js} + \beta \phi_1 g_j(a) + \phi_1 (1 + \beta) \mu a + \beta \phi_1 m_{\epsilon}^j + \sum_{k=0}^{J-1} \Psi_{ks}^j / \sum_{k=0}^{J-1} \Psi_{ks}^j
\]

where \( m_{\epsilon} \) is the mean of \( \epsilon_{jk} \).

The expected lifetime utility of the best option in the “big city” category is:

\[
E(V_1 | a, S) = \max_{j=\{0,\ldots,J-1\}} \{ \Psi_{js} + \beta \phi_1 g_j(a) + \sum_{h=0}^{J-1} \Psi_{hs}^j / \sum_{h=0}^{J-1} \Psi_{hs}^j \}
\]

The expected lifetime utility of the best option in the ”small city” category is:
\[ E(V_0 | a, S) = \max_{j \in \{0, \ldots, J_0 - 1\}} \{ \Psi_j + \beta \phi g_j \} + \sum_{h=0}^{J-1} \Psi_h s / \sum_{h=0}^{J-1} \Psi_h s + \phi(1 + \beta) \mu a + \beta \phi m_e + \beta \phi \pi_0 a \]

State S’s selection threshold for entering the labor force in a big city versus a small city is given by

\[ a^*_S = -\max_{j \in \{J_0, \ldots, J-1\}} \{ \Psi_j + \beta \phi g_j - \phi_1 \log d_{jS} \} - \max_{j \in \{0, \ldots, J_0 - 1\}} \{ \Psi_j + \beta \phi g_j - \phi_4 \log d_{jS} \} \]

where \( \Psi_j = \phi_1 \log w_j + \phi_2 \log Q_j - \phi_3 \log p_j \). \( a^*_S \) equalizes \( E(V_1 | a, S) \) and \( E(V_0 | a, S) \).

The variations in \( d_{jS} \) create variations in \( a^*_S \) among birth states: states far away from a big city tend to have relatively large selection thresholds; states in which big cities are located tend to have relatively small selection thresholds.
1.12 Appendix D: A semi-parametric approach

In the main results (Tables 1.3 and 1.4) the parameter estimates depend on normality assumption. Figure 1.8 plots the relationships between selected means and the selection probability using four common distributions. Generally, a negative correlation prevails, and the relationship is close to be linear. Therefore, instead of specifying the distribution of learning ability, I only impose a semi-parametric assumption as follows:

\[ E(a \mid a > a^*) = \delta_{10} + \delta_{11} \times P_s, \text{ and } E(a \mid a \leq a^*) = \delta_{00} + \delta_{01} \times P_s, \text{ where } \delta_{11} < 0 \text{ and } \delta_{01} < 0. \]

The results of Table 1.8 suggest that: 1) there is no static heterogeneity of learning ability (\( \mu = 0 \)); 2) the wage growth does not differ by learning ability in small cities and rural areas (\( \pi_0 = 0 \)); and 3) individuals with higher learning abilities learn more efficiently in big cities (\( \pi_1 > 0 \)).

![Figure 1.8: Approximately Linear Relationships between Selected Means and Selection Probability under Common Probability Distributions](image-url)
Table 1.8: Semi-Parametric Assumption

**Panel A. log(hourly wage), 22-26 cohort**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu \sigma_1 \delta_{11}$</td>
<td>-0.008</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>$\mu \sigma_1 \delta_{01}$</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>observations</td>
<td>35063</td>
<td>28798</td>
</tr>
</tbody>
</table>

**Panel B. log (hourly wage - city specific wage rate of the MSA of current residence)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1 \sigma_1 \delta_{11}$</td>
<td>***-0.080</td>
<td>***-0.089</td>
<td>***-0.089</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.030)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>$\pi_0 \sigma_1 \delta_{01}$</td>
<td>-0.024</td>
<td>-0.011</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.024)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>observations</td>
<td>42958</td>
<td>35262</td>
<td>31275</td>
</tr>
</tbody>
</table>

Notes:*significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. In panel A, they are clustered by MSA of current residence and birth state. In panel B, they are clustered by MSA of residence 5 years ago and birth state. In panels A and B, column 2 excludes individuals whose MSAs of current residence are coded as zero. In panel B, column 3 excludes individuals whose MSAs of current residence or residence 5 years ago are coded as zero. I place $\mu$ at zero in Panel B. The regression corresponding to column 1 in Panel B includes a dummy indicating the MSA of current residence is coded as zero. In panel B, for individuals whose MSAs of current residence 5 years ago are coded as zero, columns 1 and 2 include state-specific rural dummies to capture the average learning effect in rural areas within a state.
1.13 Appendix E: Three city categories

This section shows how to extend my empirical method to accommodate multiple city categories. As discussed in Section 1.3, knowing the ranking of the rate of dynamic return to learning ability \( \pi_j \) is necessary for estimation. In the empirical implementation, I keep the way of defining the “small city” category and further classify cities in the “big city” category with more dimensions (i.e., human capital stock and population density) besides population size.

Consider a ranking of \( \pi_j : \pi_2 > \pi_i > \pi_0 \). Let \( \pi_0 \) be the rate of dynamic return to learning ability in cities of the small category; let \( \pi_i \) be the rate of dynamic return to learning ability in cities of the medium category; let \( \pi_2 \) be the rate of dynamic return to learning ability in cities of the large category.

I follow the baseline model setup as shown in Appendix A. Two selection thresholds exist in the case of three city categories, denoted by \( a_i^* \) and \( a_0^* \). The individual with \( a \in (a_i^*, +\infty) \) chooses to enter the labor force in cities of the large category; the individual with \( a \in (a_0^*, a_i^*) \) chooses to enter the labor force in cities of the medium category; the individual with \( a \in (-\infty, a_0^*) \) chooses to enter the labor force in cities of the small category. The selection thresholds in equilibrium have the following function forms:

\[
a_i^* = -\frac{(\Psi_2 - \Psi_1) + \beta \phi (g_2 - g_1)}{\beta \phi (\pi_2 - \pi_i)}
\]
Let $T_2$ indicate whether the individual enters the labor force in a city of the large category, and $T_1$ represent whether the individual enters the labor force in a city of the medium category.

Given $j$ as the location in period one, the average wage rate in period one is:

$$E(\log w_{j,t=1} | j) = \log w_j + \mu E(a | a > a_i^*) \times T_2 + \mu E(a | a < a_i^*) \times T_1 + \mu E(a | a < a_0^*) \times (1 - T_1 - T_2)$$

Given $j$ as the location in period one and $k$ as the location in period two, the average wage rate in period two is:

$$E(\log w_{jk,t=2} | j, k) = \log w_k + g_j + (\mu + \pi_2)E(a | a > a_i^*) \times T_2 + (\mu + \pi_1)E(a | a_0^* < a < a_i^*) \times T_1 + (\mu + \pi_0)E(a | a < a_0^*) \times (1 - T_1 - T_2)$$

Assume that learning ability is normally distributed with mean at zero and variance at $\sigma_a^2$.

The selected means, $E(a | a > a_i^*)$, $E(a | a_0^* < a < a_i^*)$, and $E(a | a < a_0^*)$, can be written as the functions of selection probabilities.

$$E(a | a > a_i^*) = \sigma_a \frac{\phi(a_i^*)}{1 - \Phi(a_i^*)} = \sigma_a \frac{\phi(\Phi^{-1}(1 - P_2))}{P_2} = \sigma_a \lambda_2(P_2)$$

$$E(a | a_0^* < a < a_i^*) = \sigma_a \frac{\phi(a_0^*) - \phi(a_i^*)}{\Phi(a_i^*) - \Phi(a_0^*)} = \sigma_a \frac{\phi(\Phi^{-1}(1 - P_1 - P_2)) - \phi(\Phi^{-1}(1 - P_2))}{P_1} = \sigma_a \lambda_3(P_1, P_2)$$

$$E(a | a < a_0^*)$$
\[ E(a \mid a < a_0^*) = -\sigma_a \frac{\phi(a_0^*)}{\Phi(a_0^*)} = -\sigma_a \frac{\phi(\Phi^{-1}(1 - P_1 - P_2))}{1 - P_1 - P_2} = -\sigma_a \lambda_0(P_1, P_2) \]

where \( P_1 \) represents the selection probability for the medium category \((P_1 = \Pr(a_0^* < a < a_1^*))\), and \( P_2 \) represents the selection probability for the large category \((P_2 = \Pr(a > a_1^*))\).

Empirically, let \( P_{2S} \) be the share of college graduates born in state \( S \) who enter the labor force in cities of the large category; let \( P_{1S} \) be the share of college graduates born in state \( S \) who enter the labor force in cities belonging to the medium category. \( T_{1} \) indicates whether individual \( i \) actually enters the labor force in a city of the large category; \( T_{1} \) indicates whether individual \( i \) actually enters the labor force in a city of the medium category. Analogous to the case of two city-size categories, the specifications for estimation are given by:

\[
\log w_{ik,t} = \tau_k + \mu \sigma_a \times \lambda_2(P_{2S}) \times T_{1} + \mu \sigma_a \times \lambda_1(P_{1S}, P_{2S}) \times T_{1}
\]

(6a)

\[
\Delta \log w_{ij} = \gamma_1 Z_{1j} + \gamma_2 Z_{2j} + \gamma_3 Z_{3j} + (\mu + \pi_2) \sigma_a \times \lambda_2(P_{2S}) \times T_{1} + (\mu + \pi_1) \sigma_a \times \lambda_1(P_{1S}, P_{2S}) \times T_{1}
\]

(6b)
where
\[ \hat{\lambda}_0(P_{1s}, P_{2s}) = \phi\left(\Phi^{-1}(1-P_{1s} - P_{2s})\right)/ (1-P_{1s} - P_{2s}) \],
\[ \hat{\lambda}_1(P_{1s}, P_{2s}) = (\phi\left(\Phi^{-1}(1-P_{1s} - P_{2s})\right) - \phi\left(\Phi^{-1}(1-P_{2s})\right) )/ P_{1s} \], and
\[ \hat{\lambda}_2(P_{2s}) = \phi\left(\Phi^{-1}(1-P_{2s})\right) / P_{2s} . \]

I put MSAs with less than 1.5 million people and rural areas into the small category and experiment with three ways of grouping cities with population size above 1.5 million. I first rank these cities separately by population size, the share of college graduates in the labor force, and population density. I then pick up a cutoff for each dimension to have about one half of the sample enter the labor force in each subcategory. In total, 10, 13, and 14 of the 36 cities are in the large category when ranked by population size, the share of college graduates in the labor force, and population density, respectively.

Table 1.9 reports the results. Estimates in Panel A suggest that the static heterogeneity of learning ability remains negligible regardless of how the cities are grouped. The point estimate of \( \pi_2 \sigma_a \) (0.018 with t at 1.92) has a larger magnitude than that of \( \pi_1 \sigma_a \) (0.011 with t at 0.74) when using the human-capital-stock-based cutoff (column 2 in Panel B), while the estimate of \( \pi_0 \sigma_a \) is insignificant with a small magnitude (-0.009 with t at -0.71). However, the cutoffs based on population size (column 1 in Table 1.9) and population density (column 3 in Table 1.9) do not produce the estimates that are consistent with the ranking of \( \pi_j \). Compared to Table 1.4, the coefficients of MSA-level characteristics remain stable. The human-capital-stock-based cutoff seems to pick up cities with relatively larger rates of dynamic return to learning ability from the “big city” category. It implies that for college graduates, learning ability is rewarded more highly in big cities.
with larger sizes of human capital stock such as Washington, DC, Boston, Chicago, Minneapolis, and San Francisco.
### Table 1.9: Three City Categories

#### Panel A. log(hourly wage), 22-26 cohort

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_2(P_{25}) \times T2 )</td>
<td>0.011</td>
<td>0.006</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.014)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>( \lambda_1(P_{15}, P_{25}) \times T1 )</td>
<td>-0.017</td>
<td>-0.010</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>( \lambda_0(P_{15}, P_{25}) \times (1-T1-T2) )</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>observations</td>
<td>35063</td>
<td>35063</td>
<td>35063</td>
</tr>
</tbody>
</table>

#### Panel B. Wage growth \( \Delta \log w_{it}, 27-31 \)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(population size)</td>
<td>0.014**</td>
<td>0.019***</td>
<td>0.022***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>share college graduates in labor force</td>
<td>0.014***</td>
<td>0.010**</td>
<td>0.014***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>population density</td>
<td>0.007***</td>
<td>0.008***</td>
<td>0.007**</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>( \lambda_2(P_{25}) \times T2 )</td>
<td>0.021**</td>
<td>0.018*</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>( \lambda_1(P_{15}, P_{25}) \times T1 )</td>
<td>0.025*</td>
<td>0.011</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>( \lambda_0(P_{15}, P_{25}) \times (1-T1-T2) )</td>
<td>-0.008</td>
<td>-0.009</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>observations</td>
<td>35262</td>
<td>35262</td>
<td>35262</td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. In panel A, they are clustered by MSA of current residence and birth state. In panel B, they are clustered by MSA of residence 5 years ago and birth state. T1 means the individual enters the labor force in a city of the medium category. T2 means the individual enters the labor force in a city of the large category. The coefficients in Panel A all correspond to \( \mu \sigma_a \). In panel B, the coefficient of \( \lambda_2(P_{25}) \times T2 \) corresponds to \( \pi_2 \sigma_a \); the coefficient of \( \lambda_1(P_{15}, P_{25}) \times T1 \) corresponds to \( \pi_1 \sigma_a \); the coefficient of \( \lambda_0(P_{15}, P_{25}) \times (1-T1-T2) \) corresponds to \( \pi_0 \sigma_a \). For all the columns, I use the subsample that excludes individuals whose MSAs of current residence are coded as zero.
2. Smart City, Life-cycle Migration and Falling Mobility since the 1980s

2.1 Introduction

Rosenbloom & Sundstrom (2004), Molly, Smith & Wozniak (2011), and Kaplan & Schulhofer-Wohl (2003) among others have documented a significant decline in internal migration in the U.S. since the 1980s, which is a noticeable departure from the long-run trend since 1900. For example, the five-year cross-MSA migration rate decreased from 12 percent in 1980 to 11.4 percent in 2000, and the one-year cross-state migration rate dropped from about 3 percent in 1980 to 1.5 percent in 2010. \(^1\) It is less widely recognized that this falling migration rate is largely driven by the decline in migration from relatively large locations to relatively small ones among people in their post labor force entry period. Since 1979, a positive relationship between wage growth and city size has also become stronger. In 1979, relative to rural areas, the wage growth premium in MSAs with population sizes between 0.1 and 1.5 million was 3.6 percent, and the wage growth premium in MSAs of more than 1.5 million people was 6.5 percent. These two numbers have risen to 6.7 percent and 12.7 percent by 1999.

\(^1\) Molly, Smith, and Wozniak’s estimates are from decennial U.S. census. The five-year cross-MSA migration rate is measured by the fraction of the population that has moved within the past five years across MSA boundaries. The interstate migration rate from Kaplan and Schulhofer-Wohl (2013) is the fraction of U.S. residents at least one year old who lived in a different state one year ago calculated using the Annual Social and Economic Supplement to the Current Population Survey (March CPS).
In this paper, I examine the relocation decisions after labor force entry and propose a novel explanation for the falling migration rates since the 1980s. I suggest that a typical life-cycle migration pattern, together with the relative increase in wage growth gain in larger locations, partially contributes to the falling mobility. Young workers enter the labor force in big cities where they would be able to accumulate human capital more rapidly, reflected as greater wage growth (Glaeser and Mare, 2001; Gould, 2007; Baum-Snow and Pavan, 2011; De la Roca and Puga, 2012; Wang, 2013). After labor force entry, workers start to leave these expensive big cities for smaller locations, and the wage growth gain is portable for workers switching locations. A relative increase in wage growth gain at the location for labor force entry delays this type of relocation, causing the drop of migration from larger to smaller locations.

Following Wang (2013), I measure the wage growth differentials of American MSAs over time using two adjacent age cohorts (22-26 and 27-31) from the corresponding census. Estimates show that the wage growth differentials have been expanding in MSAs with large population sizes since 1979. By pooling the individual-level data from the 1980, 1990 and 2000 censuses, I show that the increase in wage growth gain in relatively large MSAs has induced people in 1990/2000 to stay in these more populated locations longer than their 1980 counterparts. For the age cohort 27-35 in my sample, the migration rate falls by 5.59 percentage points between 1980 and 1990 and by 6.74 percentage points between 1980 and 2000. The changes in individual characteristics generate a decline of about 1.57 percentage points and 1.45 percentage points for each period respectively. When holding the individual characteristics constant at their 1980 distribution, the changes in wage growth differentials of MSAs for labor force entry alone
generate the declines in mobility rates of 4.15 percentage points and 2.20 percentage points during the 1980-1990 and 1980-2000 time periods, which are about 74% and 33% of the actual declines in inter-MSA migration rates of these two time periods.

The existing literature has ruled out many popular explanations for this falling mobility such as changes in demographics (e.g., age, education, marital status), income, household labor force participation, occupation, industry, or homeownership (Molly, Smith & Wozniak, 2011; Kaplan & Schulhofer, 2013). In addition, Kaplan & Schulhofer (2013) discuss the decline in the geographic specificity of returns to occupations and the increase in workers’ ability to learn about local amenities in other locations due to better information technology and falling travel costs as potential causes of this decline in mobility. They show the evidence that the distribution of occupations and industries has been less geographically segregated around the country. Workers, hence, need not move to a particular place to maximize the return on their idiosyncratic abilities. They also suggest that the decrease in the cost of information disproportionally affects younger workers, causing a larger drop in their migration rate relative to older workers.² Using the Census Public Use Microdata 5 percent samples from 1980, 1990 and 2000, I plot the age profiles of five-year interstate and inter-MSA migration rates in Figure 2.1. The migration rate for each age group is measured by the fraction of people who moved across state/MSA boundaries within the past five years. The sample is restricted to native-born white men with a bachelor’s degree and a full-time full-year job who lived in the contiguous United States as of five years ago. Because of my sample choice, the

² Using a sample of working-age adults with college education from March CPS micro data, Kaplan and Schulhofer-Wohl show that since 1991, the interstate migration rate has fallen at all ages and this decline is larger for the young.
migration of people aged between 22 and 26 mainly consists of relocation from colleges/universities to places where their first jobs are located; migration in late 20s and early 30s is related to the relocation from where people first entered the labor market. In contrast to the findings of Kaplan and Schulhofer-Wohl (2013), people aged 22-26 become more mobile in more recent years, and the drop in migration rate appears only for people aged above 27. Figure 2.2 decomposes the inter-MSA migration rates of people aged 22-35 shown in Figure 2.1 by moving direction. I classify the origins and destinations into three categories: non-MSA, MSA of less than 1.5 million people in 2000, and MSA of more than 1.5 million people in 2000. For each age group, Panel A presents the fraction of people who moved to more populated locations (“move upwards”), Panel B shows the fraction of people who moved to less populated locations (“move downwards”), and Panel C plots the fraction of people who moved within each size category. At labor force entry, the majority of the inter-MSA movements consist of migration flows from less populated locations to more populated ones, and its scale has been increasing monotonically over time. This trend highlights that larger locations have become increasingly attractive to first-time job seekers. If the falling mobility is due to the less geographically segregated occupations and industries as suggested by Kaplan & Schulhofer-Wohl (2013), then we should observe falling migration rates in all charts of Panel A. However, Figure 2.2 suggests that the five-year inter-MSA migration rate actually goes up in Panel A, and the decline in mobility for people aged 27-35 shown in

---

3 For instance, the average interstate migration rate of people aged 27-35 fell from 29.8 percent in 1980 to 26.7 in 1990 and to 24.8 percent in 2000; the inter-MSA migration fell from 45 percent in 1980 to 41.2 percent in 1990 and to 40.6 percent in 2000.
Figure 2.1 is mainly driven by the drop in migration from MSAs to non-MSAs (the central and right charts in Panel B).

My paper also relates to other studies in two areas. First, the patterns of life-cycle migration have been investigated by several recent studies. By examining the migration histories of individuals between 1995 and 2000 using the 2000 U.S. census, Chen and Rosenthal (2008) explore the relationships between people’s life-cycle migration patterns and cities’ static attributes such as consumer amenities and business environments. They find that the young move towards places with a higher quality of business environment, and the old move towards places with highly valued bundles of consumer amenities. My paper considers the impacts of the dynamic attributes of cities (i.e., wage growth differentials) and focuses on relocation decisions of workers after labor force entry until mid-30s, using the census micro data in 1980, 1990 and 2000. Additionally, the type of life-cycle migration emphasized in my paper is consistent with one of various patterns of movements between big and small cities over the life cycle discussed by De la Roca, Ottaviano, and Puga (2012): workers who locate in big cities when “junior” choose to relocate to small cities when “senior” if the living cost gap between big and small cities is large, and the return to experience acquired in big cities is low. Second, evidence of the evolution of spatial wage growth disparities in cross-sections of cities found in my paper is consistent with the changes in patterns of wage structure in the U.S. labor markets documented in the literature. The relative increase in wage growth gain in larger locations found in my paper may be explained by the increasing prices of experience and skills acquired in big cities over time. Bacolod, Blum & Strange (2008) suggest that big cities enhance thinking and social interactions and thus facilitate the acquisition of cognitive
and people skills (“soft” skills). The improvements in information technology as well as the financial deregulations since 1980 may raise the demand for those “soft” skills, thereby pushing up their prices (Autor, Katz & Kearney, 2008; Philippon & Reshef, 2012). In addition, the possible increase in returns to the unobserved skills acquired in big cities since 1979 may partially contribute to the sharp increase of wage inequality in big cities since then (Baum-Snow & Pavan, 2012). Furthermore, most of the wage growth differential across city size dimension had been formed during the 1980s, and overall city size wage growth difference continued growing from 1989 to 1999 but at a much slower pace than in the 1980s. This trend may suggest that the increase in demand for the unobserved “soft” skills was less rapid in the 1990s than in the 1980s, which is parallel to the changes in the prices of observed skills over time (Autor, Katz & Kearney, 2008).

The rest of the paper is structured as follows. Section 2.2 describes the data and presents relevant patterns. Section 2.3 lays out the model. Section 2.4 shows the empirical results and the discussions. Section 2.5 concludes.

2.2 Data and Patterns

2.2.1 Data

The primary data sources for analysis are the Census Public Use Microdata 5 percent samples from 1980, 1999 and 2000 (Ruggles, et al. 2010). I use metropolitan statistical areas (MSA) to define cities. I define inter-MSA migration as a move between MSAs, between an MSA and a non-MSA, or between two non-MSAs located in different states.4

---

4 The census measures migration with a retrospective question: where did this person live 5 years ago?
There are 231 common metropolitan areas across these three censuses. People who resided in the rest of the country are assigned to the rural areas within the corresponding state.5 (Please see Appendix A for more details.) I construct information on employment, wages, mobility history (current residence and residence five years ago), and other individual characteristics (i.e., homeownership status, real hourly wages in 1984 prices, marital status, family size, whether the respondent had a child of school age five years ago, wife’s characteristics such as age, race, education attainment and employment status), for a sample of white men, aged 22-35 on December 31, 1999, with a bachelor’s degree.6 I limit my analysis to those who reported working at least 40 weeks, 35 usual hours per week and who earned at least 75 percent of the federal minimum wage in each year.7 Using the wage and migration information of two adjacent age cohorts (22-26 and 27-31) from each census, I estimate the evolution of wage growth differentials across 218 MSAs in 1980, 1990 and 2000. To show the correlation between the decline in migration rate and the relative increase of wage growth gain in the MSA where people first entered the labor force, I restrict the sample to people who lived in one of the 218 MSAs in 1975, 1985 and 1995, respectively. The empirical analysis focuses on people aged 27-35, for

5 For current analysis, I use the contemporaneous MSA boundaries. One concern is whether the drop in inter-MSA migration rates between 1980 and 2000 is due to the changes in MSA boundaries. In Figure 2.1, the trend of the age profile of interstate migration in the left panel mirrors the one of inter-MSA migration in the right panel, suggesting that this decline is not likely to be driven by the redefinition of MSA boundaries, given that the state boundaries remain stable over time. The next step of this project is to construct 1999 definition county based metropolitan area geography throughout three censuses.

6 The sample from the 1980 Census includes people with four years or above of college. The samples from the 1990 and 2000 include people who reported their educational attainment as “Bachelor’s degree.” I exclude workers with post-college education (those with graduate and professional degrees).

7 I use hourly wage rates from the age cohorts 22-26 and 27-31 to measure the wage growth differentials of American cities. Baum-Snow and Neal (2009) demonstrate that there exist significant measurement errors in hourly wages for part-time and part-year workers in the census. In addition, I use white men only to limit the possibility that changes in discrimination and patterns of labor market attachment for women and non-whites influence my estimates. I also exclude observations for those who worked in sectors of agriculture, fishing, forestry, hunting, and mining, who reported military history, or who were born outside of the contiguous United States.
whom the relocation decision is whether to leave locations where they first entered the labor force. (Figure 2.6 shows that the age profiles of interstate and inter-MSA migration rates of this subsample are similar to the full sample.)

The MSA-level variables such as population size and the share of college graduates in the labor force are calculated as aggregates from each contemporaneous census. An MSA’s population density is derived by dividing the contemporaneous population size by the 2000 MSA’s land area. Numbers for calculating population densities are from the U.S. Census Bureau. Climate variables such as precipitation, heating-degree days, cooling-degree days and January temperature are from the 2000 County and City Data Book provided by the U.S. Census Bureau. Coastal proximity (which is measured by the line distance to the nearest water including ocean and the Great Lakes) is calculated by the ArcGIS software.8

2.2.2 Spatial disparities in wage growth


---

8 Climate variables are from normal values for the 30-year period, 1961-1990. January temperature is the average daily temperature in January. One heating degree day is accumulated for each whole degree that the mean daily temperature is below 65 degree Fahrenheit. One cooling degree day is accumulated for each whole degree that the mean daily temperature is above 65 degrees Fahrenheit.
rate of the older cohort (27-31), \( \log w_{jk,i=2} \), consists of two components: the wage level component from the MSA of current residence \( k \) (five years after labor force entry) and the wage growth component from the MSA \( j \) for labor force entry. I estimate the wage level component from the earnings of the younger cohort (22-26), denoted as \( \hat{\log} w_k \), assuming that ability distribution is fixed between two age cohorts. Among the older cohort, the wage growth of individual \( i \) who worked in MSA \( j \) when first entering the labor market and currently lives in MSA \( k \), \( \Delta \log w_{ijk} \), is calculated as follows:

\[
\Delta \log w_{ijk} = \log w_{ijk,i=2} - \hat{\log} w_k
\]

Intuitively, after controlling for the wage level differences in cities of current residence, variation in the earnings of the second period come from the wage growth differentials of cities at labor force entry. In Figure 2.3, I collapse this individual wage growth measure by the size category of locations at labor force entry (non-MSAs, MSAs of less than 1.5 million people in 2000, and MSAs of more than 1.5 million people in 2000). I standardize \( \Delta \log w_{ijk} \) so that people in the “non-MSA” category have zero average wage growth. Relative to non-MSAs, the wage growth gain in MSAs have been largely increasing during the 1979-1999 time period. Between 1979 and 1999, relative to rural areas, the wage growth premium increased by 86 percent in MSAs with a population size between 0.1 and 1.5 million and by 95 percent for MSAs of more than 1.5 million people. Furthermore, overall city size wage growth difference continued growing from 1989 to 1999 but at a much slower pace than in the 1980s. Section 2.4.1 presents the details of estimation and shows that the increase in spatial disparities of wage growth mainly comes
from the city size dimension among a set of MSA-level characteristics. Based on these estimates, in Section 2.4.2, I provide evidence that compared with the 1980 counterparts, people in 1990/2000 delayed the timing of moving away from MSAs which had experienced greater increases in the wage growth differential.

2.2.3 Migration decisions at and after labor force entry

In addition to the decline in migration from relatively large locations to relatively small ones, the increase in wage growth gain in MSAs relative to non-MSAs, and in large MSAs relative to small MSAs may also raise the migration from non-MSAs to MSAs as well as from small MSAs to large MSAs after labor force entry. Panel A in Figure 2.2 shows that, at labor force entry (22-26), a significant increase in “upwards” migration occurred between 1980 and 2000. After labor force entry, we also observe a moderate increase in migration from non-MSAs to MSAs (the central and right chart in Panel A of Figure 2.2). Given the increase in the dynamic benefits of agglomeration economies, young workers find larger locations more attractive when deciding where to start their first jobs. In addition, they tend to stay in those larger locations longer after labor force entry. In this paper, I discuss the relocation decisions from relatively large to small locations after labor force entry to explain the falling internal mobility since the 1980s.9

9 For current analysis, I restrict my sample to include people aged 27-35 who resided in 218 MSAs five years ago, for which I have the estimates of the wage growth differential. For this subsample, Figure 2.7 plots their age profiles of inter-MSA migration rates by moving direction similar to Figure 2.2. While the migration rates from small MSAs (MSAs of less than 1.5 million people in 2000) to large MSAs (MSAs of more than 1.5 million people in 2000) slightly increase between 1980 and 2000, the change in inter-MSA migration for this subgroup is largely dominated by the decline in migration from MSAs to non-MSAs.
2.2.4 Other individual characteristics

For the sample of people aged 27-35, Table 2.1 presents the summary statistics of individual characteristics from three censuses. The top panel reports the distribution across one-digit industry sectors in each year. Over the past several decades, the service sector, especially the business service sector (banking and credit, security and commodity brokerage and investment companies, insurance, real estate, real estate-insurance-law offices, advertising, accounting, auditing, and bookkeeping services), has expanded while the manufacturing sector, the professional and related sector, and the public administration sector have declined. The share of home owners remained stable during this period of time. Real wages in 1984 prices, on average, have been increasing steadily over time. In addition, family characteristics have been changing since the 1980s. Family sizes and the share of married men have been declining over time. For married men (with spouse present), the share of having an employed wife has increased from about 58 percent to nearly 71 percent, which highlights the period of great growth in dual career households (Costa and Kahn, 2000). Some family characteristics are likely to increase the costs of relocating. For example, larger MSAs may provide more job opportunities than smaller locations, thereby keeping dual career couples from moving away. Furthermore, family decisions and relocation decisions may be joint choices. For instance, people may choose to delay marriage and childbearing to stay in big cities longer. I show the robustness of my main results by including those individual characteristics in the regression specifications.
2.3 Model

This model is to show how an increase in the wage growth gain in the location at labor force entry affects the timing of relocation afterwards. Each individual has a finite continuous life span $T$, $T > 0$. There are two types of locations in the economy: big and small. People initially live in the big location. They choose the timing of leaving the big location, $R$, $R \in (0, T]$. Once the individual leaves, he does not return. Wage growth is greater in the big location than in the small location, and this wage growth gain is portable for workers who relocate. Living in the big location generates disutility for workers (e.g., congestion, pollution, crimes, etc.). Given the finite lifespan, the decision when to relocate is a trade-off between the wage growth gain and the disutility from the big location.

The instantaneous utility is given by

$$
(1) \quad u(c_t) = \begin{cases} 
\ln c_t - \gamma, & \text{big} \\
\ln c_t, & \text{small}
\end{cases}
$$

where $c_t$ represents the amount of consumption in period $t$, and $\gamma$ is the disutility of living in the big location.

The wage profile is given by

$$
(2) \quad w_t = \begin{cases} 
w_0 + g \times t, & \text{big} \\
w_0 + g \times R, & \text{small}
\end{cases}
$$
where $w_0$ represents the entry wage level, and $g$ stands for the wage growth rate in the big location. $w_0 > 0$, and $g > 0$. The wage growth gain accumulated in the big location ($g \times R$) is portable between locations.

The individual chooses a consumption path, $\{c_t\}$, and the timing of leaving the big location, $R$, to maximize the lifetime utility

$$V = \int_{t=0}^{T} u(c_t) e^{-\theta t} dt$$

subject to the lifetime budget constraint

$$\int_{t=0}^{T} c_t e^{-\gamma t} dt = \int_{t=0}^{T} w_t e^{-\gamma t} dt,$$

where $\theta$ is the discount rate, and $r$ is the interest rate. For simplicity, I assume that $\theta = r = 0$.

Solving the dynamic optimization, I have

$$c^*_t = \frac{w_0 T + gRT - \frac{g}{2} R^2}{T}.$$

Plugging $c^*_t$ into (3), I rewrite the lifetime utility as a function of $R$,

$$V(R; T, w_1, w_0, \gamma) = T \ln(w_0 T + gRT - \frac{g}{2} R^2) - \gamma R - T \ln T.$$

The individual chooses $R^*$ to maximize his lifetime utility.
**Proposition:** Holding other things constant, a rise in the wage growth gain in the big location delays the timing of relocation ($\partial R^*/\partial g > 0$), and this impact is larger for younger workers ($\partial(\partial R^*/\partial g)/\partial T \geq 0$).

Proof: Please see Appendix B.

### 2.4 Empirical Results and Discussions

#### 2.4.1 Revolution of wage growth differentials across U.S. Cities, 1980-2000

The patterns from the raw data shown in Section 2.2.2 suggest that the wage growth gain in relatively large locations has been increasing since the 1980s. For the purpose of estimating the wage growth differentials of MSAs, I assume that, in each census year $t$, the wage growth differential of MSA $j$ is associated with variations in a set of MSA characteristics, denoted as $Z_{jt}$, which includes both human and natural amenity variables. Human amenity variables are log population size, population density and the share of college graduates in the labor force. Human capital accumulation is more efficient in larger and denser cities, and young workers learn more efficiently when surrounded by other highly educated workers. Besides human endowments, variation in natural amenities, such as moderate temperatures, could also affect the wage growth differential. Unfavorable climate may be bad for social activities important for knowledge spillovers, thereby inhibiting wage growth.

To construct the wage growth differential of MSA $j$ in year $t$ ($g_{jt}$), I regress the individual wage growth measure ($\Delta \log w_{ijk}$) on $Z_{jt}$ and a set of individual-level
characteristics. Table 2.2 reports the estimates for the coefficient of $Z_{jt}$, denoted as $\gamma_t$.

Human amenity variables such as the logarithm of MSA population size, population density and the share of college graduates in the labor force are calculated from contemporaneous MSAs and standardized to have mean zero and standard deviation of one. Natural amenity variables (i.e., precipitation, coastal proximity, heating-degree days, cooling-degree days, and January temperature) are constant over time. The first salient trend is that the wage growth differential along the city size dimension was absent initially in 1979 and has been increasing over decades since then. The estimate for the coefficient of log MSA population size increases from an insignificant 0.5 percent (column 1) in 1980 to a significant 1.4 percent in 1990 (column 3) and a significant 1.7 percent in 2000 (column 5). The relative increase in wage growth gain in larger locations may be explained by the increasing prices of experience and skills acquired in big cities over time. For instance, big cities enhance the acquisition of cognitive and people skills ("soft" skills), and the improvements in information technology as well as the financial deregulations since 1979 may help raise the demand for those "soft" skills, pushing up their prices (Autor, Katz & Kearney, 2008; Philippon & Reshef, 2012). Additionally, the possible increase in returns to the unobserved skills acquired in big cities since 1980 may partially contribute to the sharp increase of wage inequality in big cities since then (Baum-Snow & Pavan, 2012). Most of the wage growth differential across city size dimension had been formed during the 1980s, and overall city size wage growth

---

10 This set of individual controls include age, age squared, marital status, one-digit industry dummies, a city-size dummy indicating for MSAs of more than 1.5 million people in which people first entered the labor force, the birth state share of college graduates in MSAs of more than 1.5 million people at labor force entry and the interaction between the city-size dummy and this birth-state share. Please see details in Wang (2013).
difference continued growing from 1989 to 1999 but at a much slower pace than in the 1980s. This trend suggests that the increase in demand for the unobserved “soft” skills were less rapid in the 1990s than from the 1980s, which is parallel to the changes in the prices of observed skills over time found in the existing literature (Autor, Katz & Kearney, 2008). 11

Wage growth is also greater in MSAs where the labor force is more highly educated, and the magnitude of this correlation remains stable across decades (about 1.2-1.5 percent). In addition, the correlation between MSA population density and wage growth gain remains roughly stable with a small dip in the 1990s.

In 1979, workers in cities with either hot summers (cooling-degree days) or cold winters (heating-degree days/average January temperature) have relatively lower wage growth (columns 1-2). Unfavorable climate appears to have inhibited wage growth. Since then, the importance of climate has been declining in terms of both significance and magnitude, which may be resulting from the presence of modern, inexpensive air-conditioning.

While the coastal proximity is found to be important for explaining the wage-level differentials of cities as shown in Table 2.6, variation in coastal proximity always plays a negligible role in wage growth differentials across cities.

Using the estimates for each year in Table 2.3, I calculate the wage growth differential of MSA \( j \) in year \( t \): \( \hat{g}_{jt} = Z_{jt} \hat{\gamma}_t \). The main results are based on the estimates reported in columns 2, 4 and 6, and I test the robustness of the main results by using the estimates

---

11 The secular demand increases favoring more educated workers were less rapid in the 1990s and early 2000s than from the 1960s and the 1980s.
from columns 1, 3 and 5. Figure 2.4 plots the change in the share of people who moved away from MSAs where they first entered the labor market against the change in $\hat{g}_{jt}$ for the 1980-1990 and 1980-2000 periods, separately. The crosses represent MSAs of less than 1.5 million people in 2000 and the large circles are MSAs of more than 1.5 million people in 2000. I standardize the change in wage growth differential ($\Delta \hat{g}_{jt} = \hat{g}_{jt} - \hat{g}_{j1980}$, $T = 1990$ and 2000) so that the minimum value is zero for each period. In each graph of Figure 2.4, there exists a significantly negative correlation between these two changes. Generally, the fraction of people who moved away from the MSA where they entered the labor force declines as its $\Delta \hat{g}_{jt}$ increases, suggesting that after labor force entry, people in 1990/2000 are less likely to move away from the MSA that experienced a greater increase in wage growth differential between 1980 and 1990/2000 than their counterparts in 1980 were. Section 2.4.2 presents evidence more rigorous assessment of this observation using individual-level migration histories from three censuses.

2.4.2 Relocation decisions and wage growth differentials of MSAs for labor force entry

By pooling individual level data from different censuses, this section explores the correlation between people’s relocation decisions after labor force entry and the changes in wage growth differentials of MSAs where they first entered the labor market. I restrict my sample to include people aged 27-35 who resided in one of the 218 MSAs in 1975,
1985 and 1995.\textsuperscript{12} I pool the 1980 census with two consecutive censuses separately, and run regressions under the linear probability model (LPM) specified as below respectively.

\[
\begin{align*}
\Pr(mover_{ij} = 1) &= \sum \beta_{1i} \times \text{agegrp}_{li} + \sum \beta_{2i} \times \text{agegrp}_{li} \times I(t = T) \\
&+ \sum \beta_{3i} \times \text{agegrp}_{li} \times I(t = T) \times \Delta \hat{g}_{jt} + \mathbf{X}_i \delta + \tau_j
\end{align*}
\]

(7)

where \( j \) represents the MSA of residence five years ago, \( t \) indicates the census year, \( mover_{ij} \) indicates whether individual \( i \) relocates from \( j \) within the past five years (one means yes; otherwise means no), \( \text{agegrp}_{li} \) are age dummies, \( I(t = T) \) indicates the census year (\( T = 1990, 2000 \)), and \( \Delta \hat{g}_{jt} \) represents the change in the estimated wage growth differential of MSA \( j \) between 1980 and \( T \). \( \mathbf{X}_i \) is a vector of individual characteristics including homeownership status, log real wages in 1984 prices, marital status, family size, whether this person had a child of school age five years ago, and wife’s characteristics such as age, race, education attainment and employment status. Specification (7) includes MSA fixed effects \( \tau_j \) because of the following reason. As locations for labor force entry, MSAs are heterogeneous in terms of the fraction of the first-time job seekers who move away after labor force entry (“out-migration”). For example, workers are less likely to move away from larger cities due to the thicker labor market (Papageorgiou, 2011). The MSA fixed effects, therefore, could be correlated with both the greater increase in the wage growth differential during the post-1980 period and the smaller pre-existing out-migration rates. The inclusion of the MSA fixed effects

\textsuperscript{12} The MSAs of residence five years ago for people aged 27-35 are treated as locations where their first jobs were.
allows me to examine whether the decline in out-migration rate in an MSA is associated with the increase in its wage growth differential during the corresponding period, conditional on the pre-existing out-migration rate of this MSA. Furthermore, I bootstrap the standard errors of all the regressions with the generated regressor $\Delta \hat{g}_{jt}$.

The coefficient of interest is $\beta_{3l}$, $l = 1,...,9$. Each coefficient captures the magnitude of the change in the average likelihood of relocating after labor force entry of age group $l$ as the wage growth differential increases by one unit in the MSA at labor force entry. First, the model implies that this correlation should be negative and the impact of the change in wage growth differential decreases as the life span gets shorter. Second, for MSAs with greater increase in wage growth differential, the migration rate rises for the older cohorts as people postpone the relocation until later stages of their work lives. Therefore, $\beta_{3l}$ is expected to be negative for younger workers and increasing with age. Columns 2 and 5 report the regression results of specification (7). The estimates of $\beta_{3l}$ are significantly negative for people in their late 20s. For instance, compared with the 1980 counterpart, the likelihood of moving away from the MSA where they first entered the labor force for people aged 27 is about 2.1 percent lower in 2000 as the wage growth differential of this MSA experiences an increase of one standard deviation between 1980 and 2000 ($-0.912 \times 0.023$). The magnitude and significance generally decline with age as expected. Columns 3 and 6 present the results under the Probit Model. While the magnitudes of estimates for $\beta_{3l}$ differ between LPM and Probit, the age profile of $\beta_{3l}$ remain stable across models. Columns 1 and 4 report the regression results under LPM, without controlling for individual characteristics. Compared to the results in columns 2 and 5, the exclusion of
those individual characteristics leaves the estimates almost unchanged. Using the estimates presented in columns 1, 3 and 5 in Table 2.2, Table 2.5 reports another set of estimates for $\beta_{3i}$, which show the similar patterns to Table 2.3.

One concern is that firms located in the MSA experiencing a decrease in out-migration rate may offer their employees lower wage growth given the abundant labor supply within the region. This case would be true if firms are not fully mobile. This possible correlation, therefore, leads to an overestimation of $\beta_{3i}$ in Table 2.3 and an underestimation of the impact of the change in wage growth gain on falling migration rate.

### 2.4.3 Other changes across U.S. cities, 1980-2000

Since the 1970s, the costs of urban disamenities from crime and pollution have been falling (Glaeser 1998; Kahn, 1997), which may also contribute to the decline in the relocation from relatively large to relatively small locations after labor force entry. To control for the changes in local amenities, I include the changes in wage level differentials and housing price differentials across cities in my main regressions to examine the robustness of the results reported in Table 2.3. The rationale for this treatment is that, in the framework of the spatial equilibrium, local amenities (often called the “quality-of-life” in the literature), productivity level (or wage level) and wage growth premium are capitalized into local housing prices (Roback, 1982; Albouy, 2010; Wang, 2013). Instead of directly measuring the changes in local amenities, I include the other
two price differentials (wage level differential and housing price differential) as indirect controls. The specifications are given by:

$$\Pr(mover_{ij} = 1) = \sum_l \beta_{3l} \times agegrp_{il} + \sum_l \beta_{4l} \times agegrp_{il} \times I(t = T)$$

(8) $$+ \sum_l \beta_{5l} \times agegrp_{il} \times I(t = T) \times \Delta \hat{p}_{jt} + \sum_l \beta_{6l} \times agegrp_{il} \times I(t = T) \times \Delta \hat{w}_{jt},$$

$$+ \sum_l \beta_{7l} \times agegrp_{il} \times I(t = T) \times \Delta \hat{w}_{jt} + \delta + \tau_j$$

where $\Delta \hat{p}_{jt}$ represents the change in housing price differential in MSA $j$ between 1980 and $T$, and $\Delta \hat{w}_{jt}$ is the change in wage level differential in MSA $j$ between 1980 and $T$. I construct housing price differentials across American cities for all three census years. Appendix C presents the technical details. As shown in Table 2.6, the housing price differential has been expanding in MSAs with a more educated labor force, higher population densities, more favorable climates and better access to the ocean or the Great Lakes. The wage level differentials are estimated from the earnings of the age cohort 22-26 in the corresponding census year. The wage level differential has been increasing in MSAs where the labor force is more educated and people are more densely concentrated (Table 2.6). In addition, the spatial disparities in both housing prices and local productivity first expanded between 1980 and 1990 and then declined between 1990 and 2000 along the dimension of coastal proximity. Columns 2 and 4 of Table 2.4 report the regression results of specification (8). Columns 1 and 3 copy the results from columns 2 and 5 of Table 2.3. The age profile of the estimates for $\beta_{3l}$ remains robust after including the other two local price differentials for each time period.
2.4.4 How important is the change in wage growth differentials?

Based on the estimates from the Probit model (columns 3 and 6 in Table 2.3), I evaluate the importance of changes in individual characteristics and wage growth differentials for generating the falling migration rates in the sample of white male workers with a bachelor’s degree aged 27-35. Panel A presents the actual age profiles of the inter-MSA migration rates for the 1980-1990 and 1980-2000 periods as the benchmarks. The migration rate falls by 5.59 percentage points percent between 1980 and 1990 and by 6.74 percentage points percent between 1980 and 2000. To show the role of change in individual characteristics, for the observation in each age group, I predict each individual’s probability of leaving the MSA of residence five years ago by shutting off the change in the wage growth differential in this MSA during the periods 1980-1990 and 1980-2000. In Panel B of Figure 2.5, I collapse the predicted values by age. The change in individual characteristics generates declines of about 1.57 percentage points and 1.45 percentage points percent for each period. By further allowing the change of the wage growth differential to vary by MSA at labor force entry, I predict the probability of migrating including the effects of changes in both individual characteristics and wage growth differentials of MSAs for labor force entry. Panel C collapses the new predicted

---

13 One concern is that people in year $T$, $T = 1990, 2000$, who started work in MSAs with $\Delta \hat{\gamma}_{jT}$ below certain threshold may switch to MSAs with greater increases in wage growth differential, leading to an increase in inter-MSA migration after labor force entry. This kind of migration flow is not characterized in my model in Section 3.3. As the graphs in Figure 2.2 Panel A indicate, the increase in migration from relatively small to relatively large locations is primarily driven by the migration from non-MSAs to MSAs (the central and right graphs in Panel A of Figure 2.2) instead of between small and large MSAs (the first graph in Panel A of Figure 2.2), and the magnitude of the increase is small relative to the decrease from MSAs to non-MSAs. For empirical analysis, I include only the 218 MSAs as locations for labor force entry. In Figure 2.4, the fitted lines from linear regression (dash lines) cross at zero of Y-axis at the left end of the distribution of $\Delta \hat{\gamma}_{jT}$ for these 218 MSAs during both periods. In this paper, I treat zero as the threshold for both periods.
values by age for each time period. Comparing Panels B and C with Panel A, the inclusion of the change in wage growth differential significantly generates the gap in migration rates between census years. For the 1998-2000 period, the change in wage growth differential, along with the change in individual characteristics can explain about half of the decline in inter-MSA migration of people aged 27-35; for 1980-1990, these two factors account for most of the drop in inter-MSA migration of people aged 27-35. In Panel D, I hold the individual characteristics constant at their 1980 distribution and allow for only the changes in wage growth differentials in MSAs for labor force entry. The change in the wage growth gain in cities where people first entered the labor force alone generates the declines in mobility rates of 4.15 percentage points and 2.20 percentage points during the 1980-1990 and 1980-2000 time periods, which are about 74% and 33% of the actual declines in inter-MSA migration rates of these two periods.

2.5 Conclusions

Previous studies have documented the unprecedented trend of declining internal migration rates in the U.S. since the 1980s, and they suggest that this steady decline in mobility is not related to demographics, income, employment, labor force participation, or homeownership. Over the same time period, the dynamic benefits from agglomeration economies in large cities, in the form of greater wage growth gain, have also been increasing. This paper contributes a novel explanation of the steady decline in mobility in the U.S. to the existing literature: the increase in wage growth gain in relatively large
MSAs has induced people in 1990/2000 to stay in those more populated locations longer than their 1980 counterparts, thereby causing the drop in internal mobility.

For the sample of white male workers with a bachelor’s degree, the migration rates over three consecutive decades affirm this steady drop in internal migration. However, in contrast to the findings in previous studies, the falling migration mainly comes from the decline in migration from relatively large to relatively small locations. In addition, the migration rates of young workers at labor force entry (aged 22-26), in fact, has been slightly increasing between 1980 and 2000 due to the rise in migration from small to large locations. The age groups experiencing the major decline in mobility are people aged above 27 who entered the labor force in large metropolitan areas. Individual-level migration histories from the 1980, 1990 and 2000 censuses support the correlation between the relocation decision and the change in wage growth differentials of MSAs where people first entered the labor force. The estimates show that, for people aged 27-35 in my sample, the changes in wage growth differentials across American cities between 1980 and 2000 can account for a large fraction of falling migration for people aged 27-35. By introducing a typical lifecycle migration pattern and examining the impact of the change in city attributes on people’s relocation decisions after labor force entry, this paper improves understanding of the reasons for the trends in internal migration in the U.S. over the past three decades.
### Table 2.1: Individual Characteristics, 27-35

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(construction)</td>
<td>0.036</td>
<td>0.049</td>
<td>0.041</td>
</tr>
<tr>
<td>(manufacturing)</td>
<td>0.199</td>
<td>0.221</td>
<td>0.169</td>
</tr>
<tr>
<td>(transportation and warehousing)</td>
<td>0.022</td>
<td>0.029</td>
<td>0.028</td>
</tr>
<tr>
<td>(utilities)</td>
<td>0.024</td>
<td>0.031</td>
<td>0.030</td>
</tr>
<tr>
<td>(wholesale)</td>
<td>0.057</td>
<td>0.079</td>
<td>0.057</td>
</tr>
<tr>
<td>(retail trade)</td>
<td>0.081</td>
<td>0.102</td>
<td>0.112</td>
</tr>
<tr>
<td>(business services)</td>
<td>0.156</td>
<td>0.236</td>
<td>0.294</td>
</tr>
<tr>
<td>(personal services)</td>
<td>0.011</td>
<td>0.018</td>
<td>0.019</td>
</tr>
<tr>
<td>(entertainment and recreation services)</td>
<td>0.016</td>
<td>0.029</td>
<td>0.030</td>
</tr>
<tr>
<td>(professional and related)</td>
<td>0.276</td>
<td>0.123</td>
<td>0.148</td>
</tr>
<tr>
<td>(public administration)</td>
<td>0.121</td>
<td>0.082</td>
<td>0.071</td>
</tr>
<tr>
<td><strong>Other Individual characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(home owner)</td>
<td>0.697</td>
<td>0.683</td>
<td>0.686</td>
</tr>
<tr>
<td>Hourly wage rate (1984 prices)</td>
<td>11.834</td>
<td>13.335</td>
<td>14.627</td>
</tr>
<tr>
<td></td>
<td>(5.591)</td>
<td>(7.800)</td>
<td>(11.394)</td>
</tr>
<tr>
<td>(Married with spouse present)</td>
<td>0.722</td>
<td>0.676</td>
<td>0.636</td>
</tr>
<tr>
<td>Family size</td>
<td>2.616</td>
<td>2.548</td>
<td>2.429</td>
</tr>
<tr>
<td></td>
<td>(1.329)</td>
<td>(1.363)</td>
<td>(1.344)</td>
</tr>
<tr>
<td>(At least a child of school age 5 yrs ago)</td>
<td>0.048</td>
<td>0.036</td>
<td>0.029</td>
</tr>
<tr>
<td>observations</td>
<td>44864</td>
<td>52070</td>
<td>54113</td>
</tr>
<tr>
<td><strong>Wives' characteristics if married</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wife's age</td>
<td>29.022</td>
<td>29.593</td>
<td>30.088</td>
</tr>
<tr>
<td></td>
<td>(3.339)</td>
<td>(3.774)</td>
<td>(3.751)</td>
</tr>
<tr>
<td>(Wife is white)</td>
<td>0.990</td>
<td>0.978</td>
<td>0.959</td>
</tr>
<tr>
<td>(Wife is college graduate)</td>
<td>0.527</td>
<td>0.463</td>
<td>0.538</td>
</tr>
<tr>
<td>(Wife is currently employed)</td>
<td>0.576</td>
<td>0.693</td>
<td>0.707</td>
</tr>
<tr>
<td>observations</td>
<td>23623</td>
<td>36150</td>
<td>35037</td>
</tr>
</tbody>
</table>

Notes: Standard deviations are in parentheses. The sample is restricted to people aged between 27 and 35.
Table 2.2: Wage Growth Differentials of American Cities, 1980-2000

<table>
<thead>
<tr>
<th>log (wage growth rate)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ln(population size)</strong></td>
<td>0.005</td>
<td>0.003</td>
<td>0.014**</td>
<td>0.014**</td>
<td>0.017***</td>
<td>0.018***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td><strong>share college graduates in the LF</strong></td>
<td>0.013***</td>
<td>0.012***</td>
<td>0.013***</td>
<td>0.013***</td>
<td>0.015***</td>
<td>0.014***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td><strong>population density</strong></td>
<td>0.014***</td>
<td>0.014***</td>
<td>0.008***</td>
<td>0.008***</td>
<td>0.013***</td>
<td>0.013***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td><strong>Distance to closest water</strong></td>
<td>1.054</td>
<td>-2.214</td>
<td>-0.106</td>
<td>-1.081</td>
<td>-0.021</td>
<td>0.068</td>
</tr>
<tr>
<td>(Ocean or Great Lakes)</td>
<td>(2.615)</td>
<td>(2.639)</td>
<td>(2.730)</td>
<td>(2.720)</td>
<td>(3.024)</td>
<td>(3.032)</td>
</tr>
<tr>
<td><strong>Annual precipitation</strong></td>
<td>0.655**</td>
<td>0.202</td>
<td>-0.146</td>
<td>-0.271</td>
<td>0.152</td>
<td>0.183</td>
</tr>
<tr>
<td>(1,000,000 inches)</td>
<td>(0.315)</td>
<td>(0.290)</td>
<td>(0.287)</td>
<td>(0.275)</td>
<td>(0.303)</td>
<td>(0.313)</td>
</tr>
<tr>
<td><strong>Heating-Degree Days</strong></td>
<td>-0.007***</td>
<td>-0.009***</td>
<td>0.001</td>
<td>(1,000 days)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td><strong>Cooling-Degree Days</strong></td>
<td>-0.015**</td>
<td>-0.008</td>
<td>0.004</td>
<td>(1,000 days)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td><strong>January temperature</strong></td>
<td>5.084***</td>
<td>2.182</td>
<td>1.124</td>
<td>(1,000 degrees Fahrenheit)</td>
<td>(1.713)</td>
<td>(1.622)</td>
</tr>
<tr>
<td></td>
<td>(22.651)</td>
<td>(22.024)</td>
<td>(20.943)</td>
<td>(20.654***</td>
<td>-16.240</td>
<td>-14.496</td>
</tr>
</tbody>
</table>
Table 2.3: Relocation Decisions and Relative Changes in Wage Growth Gain of MSAs for Labor Force Entry

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPM 1</td>
<td>LPM 2</td>
<td>Probit 3</td>
<td>LPM 4</td>
</tr>
<tr>
<td>I(age 27)<em>T</em> Δ\hat{g}_{jt}</td>
<td>***-0.911</td>
<td>***-0.912</td>
<td>***-2.623</td>
<td>***-1.697</td>
</tr>
<tr>
<td></td>
<td>(0.266)</td>
<td>(0.264)</td>
<td>(0.746)</td>
<td>(0.362)</td>
</tr>
<tr>
<td>I(age 28)<em>T</em> Δ\hat{g}_{jt}</td>
<td>***-0.693</td>
<td>***-0.686</td>
<td>***-2.097</td>
<td>***-2.021</td>
</tr>
<tr>
<td></td>
<td>(0.258)</td>
<td>(0.255)</td>
<td>(0.723)</td>
<td>(0.353)</td>
</tr>
<tr>
<td>I(age 29)<em>T</em> Δ\hat{g}_{jt}</td>
<td>-0.197</td>
<td>-0.172</td>
<td>-0.623</td>
<td>*-0.607</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
<td>(0.260)</td>
<td>(0.718)</td>
<td>(0.362)</td>
</tr>
<tr>
<td>I(age 30)<em>T</em> Δ\hat{g}_{jt}</td>
<td>**-0.652</td>
<td>**-0.541</td>
<td>**-1.777</td>
<td>-0.527</td>
</tr>
<tr>
<td></td>
<td>(0.276)</td>
<td>(0.274)</td>
<td>(0.831)</td>
<td>(0.389)</td>
</tr>
<tr>
<td>I(age 31)<em>T</em> Δ\hat{g}_{jt}</td>
<td>-0.015</td>
<td>-0.002</td>
<td>-0.117</td>
<td>-0.490</td>
</tr>
<tr>
<td></td>
<td>(0.293)</td>
<td>(0.293)</td>
<td>(0.770)</td>
<td>(0.382)</td>
</tr>
<tr>
<td>I(age 32)<em>T</em> Δ\hat{g}_{jt}</td>
<td>-0.179</td>
<td>-0.109</td>
<td>-0.544</td>
<td>-0.588</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.297)</td>
<td>(0.830)</td>
<td>(0.410)</td>
</tr>
<tr>
<td>I(age 33)<em>T</em> Δ\hat{g}_{jt}</td>
<td>0.182</td>
<td>0.289</td>
<td>0.552</td>
<td>***-1.165</td>
</tr>
<tr>
<td></td>
<td>(0.289)</td>
<td>(0.284)</td>
<td>(0.842)</td>
<td>(0.424)</td>
</tr>
<tr>
<td>I(age 34)<em>T</em> Δ\hat{g}_{jt}</td>
<td>0.358</td>
<td>0.393</td>
<td>0.887</td>
<td>-0.203</td>
</tr>
<tr>
<td></td>
<td>(0.312)</td>
<td>(0.311)</td>
<td>(0.846)</td>
<td>(0.424)</td>
</tr>
<tr>
<td>I(age 35)<em>T</em> Δ\hat{g}_{jt}</td>
<td>**0.627</td>
<td>**0.677</td>
<td>**1.659</td>
<td>0.481</td>
</tr>
<tr>
<td></td>
<td>(0.309)</td>
<td>(0.310)</td>
<td>(0.924)</td>
<td>(0.440)</td>
</tr>
<tr>
<td>City fixed effects (MSA-j)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>individual characteristics</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>observation</td>
<td>86853</td>
<td>86853</td>
<td>86853</td>
<td>84810</td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Standard errors (in parentheses) are calculated on the basis of 500 bootstrap replications of the sequential estimator. Δ\hat{g}_{jt} represents the difference in the wage growth differential between census year T and 1980 in MSA j. Δ\hat{g}_{jt} is standardized for each census year T so that the minimum value in each year is zero. The Dependent variable is the dummy indicating whether the individual moved away from the MSA of residence five years ago. Individual characteristics include homeownership status, log real wages in 1984 prices, marital status, family size, whether had a child of school age five years ago, and wife's characteristics such as age, race, education attainment and employment status.
## Table 2.4: Inclusion of the Changes in Additional Local Price Differentials

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: I(move=1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I(age 27)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>***-0.912</td>
<td>***-0.865</td>
</tr>
<tr>
<td></td>
<td>(0.264)</td>
<td>(0.272)</td>
</tr>
<tr>
<td>I(age 28)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>***-0.686</td>
<td>***-0.886</td>
</tr>
<tr>
<td></td>
<td>(0.255)</td>
<td>(0.273)</td>
</tr>
<tr>
<td>I(age 29)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>-0.172</td>
<td>-0.306</td>
</tr>
<tr>
<td></td>
<td>(0.260)</td>
<td>(0.281)</td>
</tr>
<tr>
<td>I(age 30)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>***-0.541</td>
<td>*-0.495</td>
</tr>
<tr>
<td></td>
<td>(0.274)</td>
<td>(0.288)</td>
</tr>
<tr>
<td>I(age 31)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>-0.002</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>(0.293)</td>
<td>(0.324)</td>
</tr>
<tr>
<td>I(age 32)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>-0.109</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.297)</td>
<td>(0.313)</td>
</tr>
<tr>
<td>I(age 33)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>0.289</td>
<td>0.369</td>
</tr>
<tr>
<td></td>
<td>(0.284)</td>
<td>(0.330)</td>
</tr>
<tr>
<td>I(age 34)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>0.393</td>
<td>0.449</td>
</tr>
<tr>
<td></td>
<td>(0.311)</td>
<td>(0.312)</td>
</tr>
<tr>
<td>I(age 35)<em>T</em>Δ( \hat{g}_{jt} )</td>
<td>**0.677</td>
<td>**0.793</td>
</tr>
<tr>
<td></td>
<td>(0.310)</td>
<td>(0.351)</td>
</tr>
</tbody>
</table>

age dummy * T*Δ\( \hat{p}_{jt} \) | N | Y | N | Y |

age dummy * T*Δ\( \hat{w}_{jt} \) | N | Y | N | Y |

City fixed effects (MSA j) | Y | Y | Y | Y |

individual characteristics | Y | Y | Y | Y |

observation | 86853 | 86853 | 84810 | 84810 |

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Standard errors (in parentheses) are calculated on the basis of 500 bootstrap replications of the sequential estimator. Δ\( \hat{g}_{jt} \) represents the difference in the wage growth differential between T and 1980. Δ\( \hat{w}_{jt} \) represents the difference in the wage level differential between T and 1980. Δ\( \hat{p}_{jt} \) is the change in the housing price differential between T and 1980. Δ\( \hat{g}_{jt} \) is standardized so that the minimum value in each year is zero. The Dependent variable is the dummy indicating whether the individual moved away from the MSA of residence five years ago. Columns 1 and 3 copy the results indicating whether the individual moved away from the MSA of residence five years ago. Columns 1 and 3 copy the results from columns 2 and 5 of Table 3.
Table 2.5: Results under the Alternative Estimates of Wage Growth Differentials

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPM 1</td>
<td>LPM 2</td>
</tr>
<tr>
<td>I(age 27)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>***-1.378</td>
<td>***-1.369</td>
</tr>
<tr>
<td></td>
<td>(0.266)</td>
<td>(0.272)</td>
</tr>
<tr>
<td>I(age 28)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>***-0.688</td>
<td>***-0.642</td>
</tr>
<tr>
<td></td>
<td>(0.258)</td>
<td>(0.246)</td>
</tr>
<tr>
<td>I(age 29)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>-0.229</td>
<td>-0.165</td>
</tr>
<tr>
<td></td>
<td>(0.267)</td>
<td>(0.270)</td>
</tr>
<tr>
<td>I(age 30)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>**-0.630</td>
<td>**-0.506</td>
</tr>
<tr>
<td></td>
<td>(0.309)</td>
<td>(0.273)</td>
</tr>
<tr>
<td>I(age 31)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>0.042</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.304)</td>
<td>(0.311)</td>
</tr>
<tr>
<td>I(age 32)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>-0.179</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>(0.323)</td>
<td>(0.305)</td>
</tr>
<tr>
<td>I(age 33)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>0.374</td>
<td>0.496</td>
</tr>
<tr>
<td></td>
<td>(0.321)</td>
<td>(0.312)</td>
</tr>
<tr>
<td>I(age 34)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>*0.556</td>
<td>*0.576</td>
</tr>
<tr>
<td></td>
<td>(0.325)</td>
<td>(0.305)</td>
</tr>
<tr>
<td>I(age 35)<em>T</em> $\Delta \hat{g}_{JT}$</td>
<td>**0.791</td>
<td>**0.861</td>
</tr>
<tr>
<td></td>
<td>(0.322)</td>
<td>(0.333)</td>
</tr>
<tr>
<td>City fixed effects (MSA j)</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>individual characteristics</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>observation</td>
<td>86853</td>
<td>86853</td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Standard errors (in parentheses) are calculated on the basis of 500 bootstrap replications of the sequential estimator. $\Delta \hat{g}_{JT}$ represents the difference in the wage growth differential between census year T and 1980. $\Delta \hat{g}_{JT}$ is calculated using the estimates from columns 1, 3 and 5 of Table 2 and standardized for each census year T so that the minimum value in each year is zero. The Dependent variable is the dummy indicating whether the individual moved away from the MSA of residence five years ago.
<table>
<thead>
<tr>
<th></th>
<th>housing price differentials</th>
<th>wage level differentials</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(population size)</td>
<td>0.042**</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>share college graduates in labor force</td>
<td>0.082***</td>
<td>0.183***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>population density</td>
<td>0.025***</td>
<td>0.107***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Distance to closest water (Ocean or Great Lake)</td>
<td>-39.966***</td>
<td>-105.320***</td>
</tr>
<tr>
<td></td>
<td>(8.096)</td>
<td>(13.279)</td>
</tr>
<tr>
<td>Annual precipitation (1,000,000 inches )</td>
<td>-8.874***</td>
<td>-9.205***</td>
</tr>
<tr>
<td></td>
<td>(1.017)</td>
<td>(1.697)</td>
</tr>
<tr>
<td>Heating-Degree Days (1,000 days)</td>
<td>-0.068***</td>
<td>-0.127***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Cooling-Degree Days (1,000 days)</td>
<td>-0.110***</td>
<td>-0.219***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>observations</td>
<td>229</td>
<td>229</td>
</tr>
</tbody>
</table>

Notes:*significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. Regressions for housing price differentials are weighted by the number of households in each MSA. Regressions for wage level differentials are weighted by the number of observations in each MSA in the earnings regression of the younger cohort.
2.7 Figures

Figure 2.1: Age Profiles of Interstate and Inter-MSA Migration Rates

Notes: The sample is restricted to native-born white male workers who report working at least 40 weeks, 35 usual hours per week and who earn at least 75 percent of the federal minimum wage in each census year and lived in the contiguous United States five years ago. Data is from the census 5% PUMS in 1980, 1990 and 2000. The migration rate for each age group is the fraction of people who moved across state/MSA boundaries within the past five years.
**Figure 2.2: Decomposition of Inter-MSA Migration by Moving Direction**

Notes: The sample is restricted to native-born white male workers who report working at least 40 weeks, 35 usual hours per week and who earn at least 75 percent of the federal minimum wage in each census year and lived in the contiguous United States five years ago. Data is from the census 5% PUMS in 1980, 1990 and 2000. The migration rate for each age group is the fraction of people who moved across state/MSA boundaries within the past five years.
Figure 2.3: Average wage growth by Size Category, 1980-2000

Notes: I collapse the individual wage growth measure by the size category of locations for labor force entry (non-MSAs, MSAs of less than 1.5 million people and MSAs of more than 1.5 million people in 2000). I standardize the wage growth measure in each panel so that people in the “non-MSAs” category have zero average wage growth in each census year.
Figure 2.4: Changes in out-Migration Rate against Changes in Wage Growth Differential by MSA for Labor Force Entry
Figure 2.5: Actual and Predicted Age Profiles of Inter-MSA Migration Rates, 27-35

Notes: For the predictions in Panels C and D, I standardize the change in wage growth differentials of MSAs for labor force entry (\(\Delta \hat{g}_{jT}\)) so that the minimum value in each year is zero.
Figure 2.6: Age Profiles of Interstate and Inter-MSA Migration Rates, Subsample
Figure 2.7: Decomposition of Inter-MSA Migration by Moving Direction, Subsample
2.8 Appendix A: Data Appendix

Originally, I identify 272 MSAs in the 1980 census, 273 MSAs in the 1990 census, and 297 MSAs in the 2000 census. There are 231 MSAs in all three censuses. People who resided in the rest of the country are treated as living in rural areas within the corresponding state. In order to maximize the observations in MSAs, I consolidate some “extra MSAs” in each census to those 231 MSAs. For example, there is no observation in Athens, GA in 1990. Geographically, Athens, GA is directly adjacent to Atlanta, GA, which is present in all three censuses. I then assign the MSA code of Atlanta, GA to people who reported living in Athens, GA in the 1980 and 2000 census.
2.9 Appendix B: Proof of the Proposition

Proof: The first order condition implies that

\[(A1) \ F \triangleq \frac{1}{2} g\gamma R^* - g(\gamma + 1)TR^* + gT^2 - w_0 T\gamma = 0\]

There are two solutions for (A1), and only the smaller one is valid given that \(R^* < T\). The function form of \(R^*\) is given by

\[R^* = \frac{g\gamma T + gT - \sqrt{(\gamma^2 + 1)g^2 T^2 + 2g\gamma}}{g\gamma} < \frac{g\gamma T + gT - \sqrt{g^2\gamma^2 T^2}}{g\gamma} = \frac{T}{\gamma}.\]

I have

\[\frac{\partial F}{\partial R^*} = g\gamma R^* - g\gamma T - gT < 0\]

\[\frac{\partial F}{\partial g} = \frac{\gamma}{2} R^2 - (\gamma + 1)TR^* + T^2,\]

From equation (A1), I have \(g\left(\frac{\gamma R^2}{2} - (\gamma + 1)TR^* + T^2\right) = w_0 T\gamma > 0\). Therefore, \(\frac{\partial F}{\partial g} > 0\).

By Implicit Function Theorem,

\[\frac{\partial R^*}{\partial g} = -\frac{\frac{\partial F}{\partial g}}{\frac{\partial F}{\partial R^*}} = \frac{T^2 + \frac{\gamma}{2} R^2 - (\gamma + 1)TR^*}{g\gamma T + gT - g\gamma R^*} > 0.\]
Additionally,

\[
\frac{\partial (\partial R^* / \partial g)}{\partial T} = \frac{1}{(g\gamma T + gT - g\gamma R^*)^2} \left( (\gamma + 1)gT^2 - 2\gamma gR^*T + \frac{(\gamma + 1)\gamma gR^*}{2} \right)
\]

\[
> \frac{1}{(g\gamma T + gT - g\gamma R^*)^2} \left( (\gamma + 1)gT^2 - 2\gamma gR^*T \right)
\]

\[
> \frac{1}{(g\gamma T + gT - g\gamma R^*)^2} \left( (\gamma + 1)gT\gamma R^* - 2\gamma gR^*T \right) \quad (T > \gamma R^*)
\]

\[
= \frac{1}{(g\gamma T + gT - g\gamma R^*)^2} (\gamma - 1)^2 gTR^* \geq 0
\]
2.10 Appendix C: Construct the housing price differentials

The housing price differentials of U.S. cities are calculated separately for three census years using the logarithm reported gross rents and housing values. I include all household heads from the Census Public Use Microdata five percent samples from 1980, 1990 and 2000. The housing price differential of an MSA is found by regressing the logarithm reported gross rents and housing values on a set of covariates at the unit level and indicators for which MSA a worker lives in, using the coefficients of these MSA indicators. The covariates for the adjusted differentials are: 3 indicators for the number of housing units in the structure containing the household, 9 indicators for the number of rooms, 5 indicators for the number of bedrooms, number of rooms interacted with number of bedrooms, 2 indicators for lot size, 6 indicators for when the building was built, 2 indicators for complete plumbing and kitchen facilities, an indicator for commercial use and an indicator for condominium status. The estimation method follows Albouy (2010). A regression of housing values on housing characteristics and MSA indicator variables is first run using only owner-occupied units, weighting by household weights. A new value-adjusted weight is calculated by multiplying the household weight by the predicted value from the first regression using housing characteristics alone, controlling for MSA. A second regression is run using these new weights for all units, rented or owner-occupied, on the housing characteristics fully interacted with tenure as well as the MSA indicators (which are not interacted). The housing price differentials are taken from the MSA indicator variables in the second regression.
3. Fundamentals in China’s Housing Markets

3.1 Introduction

Since 2000, escalating residential housing prices have been a remarkable phenomenon in urban China, and the appreciation rate varies largely across major Chinese housing markets. As shown in Figure 3.1, the average sale price across 35 major Chinese cities increased by 10.53 percent annually between 2002 and 2008.¹ Cities in the east coast, such as Beijing, Fuzhou, Ningbo and Xiamen, experienced the annual appreciation of about 15 percent, while the average prices of the inland cities, such as Ha’erbin, Kunming, Shenyang and Yinchuan, increased by around 2-5 percent. How important are changes in fundamental factors, such as urban hukou population, wage income, urban land supply and construction cost, in explaining the changes in housing prices in major Chinese cities?² In which city has the housing price appreciation significantly deviated from changes in those fundamental factors?

To pursue our research questions, we evaluate how much of the variation in housing price appreciation across the 35 major Chinese cities can be explained by variations in changes of the fundamental factors. We first derive equilibrium housing price appreciation as a

¹ The 35 cities are Beijing, Changchun, Changsha, Chengdu, Chongqing, Dalian, Fuzhou, Guangzhou, Guiyang, Ha’erbin, Haikou, Hangzhou, Hefei, Huhehaote, Jinan, Kunming, Lanzhou, Nanchang, Nanjing, Nanning, Ningbo, Qingdao, Shanghai, Shenyang, Shenzhen, Shijiazhuang, Taiyuan, Tianjin, Wuhan, Wulumuqi, Xi’an, Xiamen, Xining, Yinchuan, and Zhengzhou. About half of the completed national investment for residential housing in 2008 occurred in these 35 cities.

² Household registration is called hukou in China. We use a city’s non-agricultural hukou population to approximate local urban hukou holders.
function of the fundamental factors of supply and demand. We then estimate and calibrate the coefficients of these fundamental factors, using separate data sets. Finally, based on the estimated coefficients, we calculate each city’s housing price appreciation from the observed changes in fundamental factors and compare the actual appreciation with the calculated equilibrium counterpart. The results show that for most of the cities in our sample, actual housing price appreciation can be largely explained by changes in fundamentals. Seven cities seem to have “overly high” actual appreciation rates relative to the other cities, which may suggest risks of “bubbles”. 3

To estimate the coefficients of fundamental factors in the equation of equilibrium housing price appreciation, we first apply the method proposed by Epple, Gordon, and Sieg (2010) to estimate the land-share parameter in the housing production function in urban China, using a sample of matched land sale and residential development project data. Assuming a Cobb-Douglas production function with constant-returns-to-scale technology, we estimate the land-share parameter in the housing production function to be about 0.323. Second, we estimate the lower bound of the impact of population growth among local urban hukou holders on local housing price appreciation. We find that, on average, a 1 percent population growth among local urban hukou holders caused at least a 0.54 percent increase in the sale price of residential housing. Third, using the above two estimates and borrowing income and price elasticities of housing demand from the literature (around 1.0 and -0.5 in the case of China, respectively), we infer the coefficients of wage income and construction cost.

---

3The seven cities are Beijing, Fuzhou, Hangzhou, Ningbo, Tianjin, Wuhan, and Xiamen.
Our paper builds upon the existing literature that examines the roles of fundamental factors of supply and demand in housing price appreciation and contributes new evidence from urban housing markets in China. First, many papers within this literature examine how the demand shifters such as demographics, income and credit market affect the appreciation of real estate prices (e.g., Mankiw and Weil, 1989; Ortalo-Magne and Rady, 2006; Ferreira and Gyourko, 2011). The institutional changes in the household registration system (“hukou” system) in the late 1990s and early 2000s created a massive amount of new buyers for houses in urban China, who had been once excluded from the formal urban housing markets. Since 2000, many major Chinese cities have experienced an unprecedented growth in urban hukou population; for the 35 cities, urban hukou population has increased by about 4 percent annually. In addition, the real wage income of these 35 cities grew at an annual 11 percent on average. The rapid urbanization process and the fast income growth are strong driving forces behind the booming real estate markets. Second, another set of papers address the causal impact of inelastic housing supply due to both physical and regulatory constraints on the spatial dispersion of housing price appreciation since 1970 (e.g., Glaeser, Gyourko and Saks, 2005; Saiz, 2010). In China, urban land supply is under strict government regulations. During our study period, the average land price per square meter has soared by 21 percent annually, from 467 yuan in 2002 to 1,439 yuan (2005 yuan) in 2008. It is a hotly debated policy issue whether the local governments are too slow in adjusting the land supply to meet

---

4 Mankiw and Weil (1989) find that the entry of the Baby Boom generation into its house-buying years is the major cause of the increase in real housing prices in the 1970s. Ortalo-Magne and Rady (2006) suggest that the ability of young households to afford the down payment on a starter home is a powerful driver of the housing market. Ferreira and Gyourko (2011) find that local income is the only potential demand shifter found that also had a significant change around the time that local housing booms began. They also find that the key indicators in the lending market such as the share of subprime lending do lag the beginning of the boom.
demand changes. For the purpose of investigating how much the land supply constraint contributes to the housing price appreciation, we estimate the land share in the housing production function in China. To our knowledge, we are the first in the literature to estimate the housing production function for urban China using micro-level land and housing data. 5 Third, our paper proposes a method to evaluate the housing price appreciation driven by the fundamental factors from both the demand side and the supply side for each city. This method provides us a standard to assess whether a city’s housing price appreciation has gone beyond its equilibrium counterpart driven by changes in several main fundamental factors.

The rest of the paper proceeds as follows. Section 3.2 develops a simple theoretical framework to derive the equilibrium housing price equation. Section 3.3 discusses the empirical strategy. In Section 3.4, we estimate the land-share parameter in the housing production function. In section 3.5, we estimate the population elasticity in the equilibrium housing price equation. In Section 3.6, we calculate a city’s equilibrium housing price appreciation from changes in fundamentals and compare it to the actual one. Section 3.7 concludes.

5Epple, Gordon, and Seig (2010) find the land share to be 0.144, using separate assessed land and structure values for houses in Alleghany County, PA and assuming a Cobb-Douglas production function with the constant-returns-to-scale technology. Albouy and Ehrlich (2011) find, on average, 30 percent of housing costs are due to land, which ranges from 0.13 to 0.51 across metropolitan areas in the US, accounting in sequence for non Cobb-Douglas production function, geographic and regulatory constraints, non-land input costs, and disaggregated measures of regulatory constraints.
3.2 Theoretical Framework

We use a standard supply-demand framework to understand how fundamentals determine the equilibrium residential housing prices. The purpose is to specify the equilibrium housing price appreciation as a function of the fundamental factors of demand and supply.

We first model the supply side. Consider a city which is a geographic market for urban residential houses and land. The urban residential land market and urban residential housing market are two relevant markets in the model. In the urban residential housing market, real estate developers are suppliers. As observed from the land auction data provided by Cai, Henderson, and Zhang (2010), the number of developers in any major city is large. We thus assume perfect competition among developers in a city.

Developers are price takers in both the land and housing markets. We assume that they have the same production technology which exhibits constant returns to scale. $\bar{L}$ is the total land stock for residential housing supplied by the local government. Developer $i$ uses a land parcel of size $L_i$ at rent $p_L$. This model assumes that housing is a homogeneous and perfectly divisible good denoted by $H$. Housing is produced from two factors $M$ and $L$ via a Cobb-Douglas production function,

$H = H(M, L) = AM^\alpha L^{1-\alpha}, \ A > 0, \ 0 < \alpha < 1,$

---

6 In China, urban land supply is under strict government regulations. All urban land is owned by the state. What is transacted in the market, through public auctions, is leasehold of land. There are three auction types in China’s urban land auction: English auction (called paimai in Chinese), two-stage auction (guapai), and sealed-bid auction (zhaobiao). For detailed description of these auction types, see Cai, Henderson, and Zhang (2012). City government is the monopoly supplier of land to real estate developers. As one part of the urban land supply, each year, city government converts a certain amount of rural land owned by the farmers’ collectives to urban land. The other part of the supply comes from urban re-development, which is again planned by the city government.
where \( L \) is land, \( M \) is a composite of all mobile non-land factors, and \( A \) represents housing production productivity. The price of \( M \) is denoted by \( p_m \). Developer \( i \) maximizes its profit

\[
\Pi_i = p_h AM_i^\alpha (L_i)^{1-\alpha} - p_m M_i - p_L L_i.
\]

The first-order condition gives \( M_i^* = \left(\alpha A p_h / p_m \right)^{1/(1-\alpha)} L_i \). The aggregate housing supply, therefore, is

\[
H = \sum_i (M_i^*)^\alpha (L_i)^{1-\alpha} = \left(\alpha A p_h / p_m \right)^{\alpha/(1-\alpha)} \sum_i L_i = \left(\alpha A p_h / p_m \right)^{\alpha/(1-\alpha)} L.
\]

Taking logarithm on both sides of the aggregate housing supply equation, we have:

\[
(2) \ln H = \rho_S + \ln L + \frac{\alpha}{1-\alpha} \ln p_h - \frac{\alpha}{1-\alpha} \ln p_m,
\]

where \( \rho_S = (\alpha / (1-\alpha)) \ln(\alpha A) \).

We describe the demand side by approximating the aggregate housing demand as follows:

\[
(3) \ln H = \delta_1 \ln w + \delta_2 \ln N - \delta_3 \ln p_h,
\]

where \( \delta_i > 0, \ i = 1,2,3 \). \( N \) is city’s population size, \( w \) represents city’s average income, and \( p_h \) is housing price.

From (2) and (3), we have the equilibrium housing price:
(4) \( \ln p_h = \phi_0 + \phi_1 \ln w + \phi_2 \ln N - \phi_3 \ln L + \phi_4 \ln M \),

where \( \phi_0 = -\rho \frac{\delta (1-\alpha)}{\alpha + \delta (1-\alpha)} \), \( \phi_1 = (\delta (1-\alpha)) / (\alpha + \delta (1-\alpha)) \), \( \phi_2 = (\delta (1-\alpha)) / (\alpha + \delta (1-\alpha)) \), \( \phi_3 = (1-\alpha) / (\alpha + \delta (1-\alpha)) \), and \( \phi_4 = \alpha / (\alpha + \delta (1-\alpha)) \).

Differencing (4) on both sides derives the following relationship between equilibrium housing price appreciation and the fundamental factors:

(5) \( \Delta \ln p_h = \Delta \phi_0 + \phi_1 \Delta \ln w + \phi_2 \Delta \ln N - \phi_3 \Delta \ln L + \phi_4 \Delta \ln M \).

In (5), the growth in a city’s average income (\( \Delta \ln w \)) and population size (\( \Delta \ln N \)) work as demand shifters, which positively affect the growth in the city’s equilibrium housing price; the growth in urban residential land stock (\( \Delta \ln L \)) and construction cost (\( \Delta \ln M \)) are supply shifters, which have negative and positive impacts on the city’s housing price appreciation in equilibrium, respectively. \( \Delta \phi_0 \) represents the common growth factor.

### 3.3 Empirical Strategy

To answer our research questions, the basic strategy is as follows. If we know all the coefficients in (5), for each city we can directly compare the actual housing price appreciation with the equilibrium counterpart calculated from the changes of fundamental factors. A significant positive deviation (the actual minus the calculated) may suggest a risk for “bubbles.” The conventional method for estimating these coefficients is to obtain the OLS estimates by directly regressing housing price changes on changes in the fundamental factors using the follow specification:
where \( j \) indexes city, and \( \varepsilon_j \) is the error term. However, the estimates would be biased due to the endogeneity issues. For example, local government may accommodate a positive demand shift with an increase in land supply, which leads to an upward bias for the estimate of \( \phi_3 \); high housing prices might discourage potential permanent migrants from settling down in this city, which would lead to a downward bias for the estimate of \( \phi_2 \).

Instead of directly estimating those coefficients in a single OLS regression, we estimate each coefficient separately, using both city level and residential development project data. We first estimate the land-share parameter in the housing production function \((1-\alpha)\), using a sample of matched land sale and residential development project data. Second, an instrumental variable strategy is implemented in the long-difference specification to estimate the lower bound of the impact of hukou population on housing price \( \phi_2 \).

Finally, using the above two estimates and borrowing the income and price elasticities from the literature, we infer all coefficients of fundamental determinants in (5).

### 3.4 Housing Production Function Estimation

We follow the method proposed by Epple, Gordon, and Sieg (2010) to estimate the housing production function. Like Epple, Gordon, and Sieg, we use a micro-level data set, including information on land parcel sales and their matched residential development (loupan), to estimate the housing production function. Information on land parcel sales
and their ex post residential developments is from the matched land auction and *loupan* data sets, provided by Cai, Henderson, and Zhang (2012). The land auction data set is for completed sales from the Land Bureau of China (or its branches at city level). It reports information on the area, the type of auction (two-stage auction, English auction, and sealed-bid auction), the sale price if the sale is completed, and the sale date. It also has the geo-economic characteristics of each piece of land for sale: the line distance between the land parcel and the city’s central business district (CBD) and whether within a 2.5-kilometer radius of the center of the property for sale, there is railway (including light rail and subway) or highway. In the *loupan* data set, for each development, there is an average price per square meter of floor space and the actual floor-area-ratios as of April 2009. Each residential development is matched to a land parcel by its address and the identity of its developer. The match requires that the *loupan* has new property for sale in April 2009. We have 505 matched pairs from 23 cities for the regression analysis. To implement Epple, Gordon, and Sieg’s method, for each match, we measure its price per unit of land using the land parcel’s sale price divided by the total land area and calculate the corresponding housing value per unit of land using the *loupan*’s average price per square meter multiplied by the floor-to-area-ratio. Details on data construction and summary statistics for main variables are in Appendix A and Table 3.1 respectively.

From Equation 1, we have \( h(m) = Am^\alpha \), which is the production function per unit of land, where \( h = H / L \) and \( m = M / L \). The firm’s profit per unit of land is given by

\[
\pi = p_h h(m) - p_m m - p_f.
\]

From profit maximization, we have

---

7 These are Beijing, Changchun, Chengdu, Chongqing, Dalian, Fuzhou, Guangzhou, Hangzhou, Nanchang, Nanjing, Nanning, Ningbo, Shanghai, Shenyang, Shenzhen, Suzhou, Taiyuan, Tianjin, Wuhan, Wuxi, Xi’an, Xiamen, and Zhengzhou.
(7) \( m(p_h) = (A \alpha p_h / p_M)^{\alpha/(1-\alpha)} \).

Therefore, the supply function per unit of land is

(8) \( S(p_h) = q(m(p_h)) = A(A \alpha p_h / p_M)^{\alpha/(1-\alpha)} \).

Hence, housing value per unit of land is

(9) \( v = p_h S(p_h) = A(A \alpha / p_M)^{\alpha/(1-\alpha)} p_h^{\alpha/(1-\alpha)} \).

From zero-profit condition, we have

(10) \( p_l = v - p_m m(p_h) \).

Here, \( p_l \) is the price per unit of land.

Based on (7)-(10), we have the following equation, which is used to estimate the housing production function.

(11) \( p_l = (1 - \alpha)v \).

The left-hand-side variable is the price per unit of land, and the right-hand-side variable is the housing value per unit of land. \( (1 - \alpha) \) is the land-share parameter.

Let the observed price per unit of land for each land parcel-\( loupan \) match be

\( \tilde{p}_l = p_l (1 - \tau \times auctiontype) + \varepsilon_p \), where \( p_l \) is the actual unit land cost paid by the land developer, \( \varepsilon_p \) is the measurement error, \( auctiontype \) indicates the land sale was done
through a two-stage auction (otherwise, it’s through English auction), and \( \tau > 0.8 \). Cai, Henderson, and Zhang (2012) find that land sales through two-stage auctions are more corruptible than English auctions, and their empirical results suggest that, overall, sales prices are significantly lower for two-stage auctions than English auctions. Corrupt land sales would cost land developers extra money to bribe local land officers, causing the gap between the actual unit land cost (\( p_i \)) and the observed price per unit of land (\( \tilde{p}_i \)). We assume that this gap is proportional to the actual price per unit of land (\( \tau \times p_i \)).

The value of housing per unit of land for each land parcel-loupan match may also be measured with error \( \varepsilon_v \): \( \tilde{v} = v + \varepsilon_v \), where \( v \) is the value of housing per unit of land perceived by the land developer. As Epple, Gordon, and Sieg (2010) suggest, \( \varepsilon_v \) may reflect either measurement error or productivity shocks (for example, the uncertainty about the appeal of the completed structure). Substitute both equations into (11), we have

\[
\tilde{p}_i = (1 - \alpha)\tilde{v} - (1 - \alpha)\tau \times \text{auctiontype} + \varepsilon_p - (1 - \alpha)\varepsilon_v + (1 - \alpha)\tau \times \varepsilon_v \times \text{auctiontype}.
\]

For the estimation purpose, we rewrite it as

\[
(12) \quad \tilde{p}_i = \beta_1 \tilde{v}_i + \beta_2 \varepsilon_v \times \text{auctiontype}_i + u_i,
\]

where \( i \) indexes land parcel-loupan match, \( u_i = \varepsilon_p - \beta_1 \varepsilon_v - \beta_2 \varepsilon_v \times \text{auctiontype}_i \), \( \beta_1 = (1 - \alpha) \), and \( \beta_2 = -(1 - \alpha)\tau \).

---

\( ^{8} \) Among 505 land parcels in our sample, 391 land parcels were sold through two-stage auctions, and 114 were through English auctions.
An endogeneity problem exists in the estimation of (12) because $\text{cov}(\bar{y}_i, u_i) = -\beta \text{var}(\varepsilon_i)$.

This means that the OLS estimate will be attenuated. We use the following variables to instrument for the observed housing value per unit of land: $dist_i$, which is the distance between the land parcel and the city’s CBD; $railway_i$, which represents whether a railway exists within a 2.5-kilometer radius; and $highway_i$, which represents whether a highway exists within a 2.5-kilometer radius. The auction type of land sale, represented by $auctiontype_i$, is also controlled in the first stage because properties with better unobserved characteristics are likely to be selected into corrupt land sales (Cai, Henderson, and Zhang, 2012).

The following assumptions for valid IVs are required. First, the IVs should strongly affect the observed housing value per unit of land. This assumption is likely to hold, since the property values tend to decline as commuting time rises (Zheng and Kahn, 2008) and increase for corrupt land sales (Cai, Henderson, and Zhang, 2012). Second, the geographic characteristics should be uncorrelated with the measurement errors, which is likely to be valid. Third, it is unlikely that the uncertainty about the appeal of the completed structure correlates with the proximity to the CBD or to the public transit. Lastly, the auction type of land sale is unlikely to correlate with either $\varepsilon_p$ or $\varepsilon_v$.

Results are shown in Table 3.2. Column 1 presents the first-stage results. Our instruments are strong (F-static is 34.812). Similar to the empirical results of Zheng and Kahn (2008), we find evidence that the proximity to CBD and public transit infrastructure are capitalized into real estate prices. The OLS estimate (Column 2) is smaller than the 2SLS
(Column 3), implying the existence of attenuation bias due to measurement errors. There are no constant terms in the OLS and the 2SLS specifications because $p_i$ must go to zero as $v$ goes to zero. The estimated land-share parameter $(1-\alpha)$ in the case of China is 0.323 (Column 3). As a comparison, the estimate for this parameter ranges from 0.13 to 0.51 across metropolitan areas in the U.S. (Epple, Gordon and Seig, 2010; Albouy and Ehrlich, 2011).

3.5 Population Elasticity Estimation: Long-difference Specification

In this section, we estimate the population elasticity ($\phi_2$) in the equation of equilibrium housing price. In view of the institutional background in China, we use the instrumental variables (IVs) strategy in a long-difference specification to estimate the lower bound of the impact of population growth among local urban hukou holders on housing price appreciation between 2002 and 2008.

The hukou system was introduced in 1958 to regulate the migrations of all people within the country. Between 1958 and 1978, rural-urban migration was banned. Since 1978, restrictions on the non-hukou migration have been relaxed due to the need for economic development in the coastal cities.\footnote{According to the calculation by Chan (2010), the size of the rural migrant labor force grew from about 50-60 million in the early 1990s to exceed 100 million in the early 2000s.} However, despite the looser constraints on non-hukou migration, before the late 1990s, the growth rate of local urban hukou holders (hukou migration) was controlled strictly by the state using a quota system. Since 2000, local authorities (cities and provinces) have been gaining more autonomy in determining the admission of new legal urban citizens, due to the decentralization of China’s national
management system (e.g., the fiscal system). The quota system was replaced by a conditional system. In this new system, migrants with better socio-economic characteristics have a greater chance of being granted the local urban hukou; relatives (old parents, young children, and spouses) of current local urban hukou holders can be granted the local urban hukou. This change has facilitated many old and new “non-hukou migrants” to transform themselves into hukou migrants, either from the same province or from the other provinces. These new legal urban citizens become new buyers in local formal housing markets and push up local housing prices. More details on the trends of population growth among urban hukou holders and the institutional changes since 1949 are discussed in Appendix B.

Data used here are from multiple sources. Housing price data for city-level regressions are from National Statistical Yearbooks, which report the average sale prices of residential housing for 35 major Chinese cities, including newly-built and representative second-hand transactions from 2002-2008. One drawback of this dataset is that housing quality is not controlled for. But, to the best of our knowledge, this is the only housing price measure that covers most of the major Chinese cities with a decent time span.

We measure a city’s population size as the total population of local urban hukou holders, which is reported in National Urban yearbooks. Hukou classifies resident into “local” (hukou is in the place of residence) versus “migrant” (hukou is in somewhere else other than the place of residence), as well as “rural” versus “urban.” The percentage of local urban hukou holders living in private-rental housing was below ten percent in 2000 (Logan, Fang, and Zhang, 2005). Rural migrants live either in the outskirts of the city
proper where low-quality and cheap rental housing is available, or in dorms or temporary housing offered by their employees. Therefore, we assume that only local urban hukou holders can access the formal residential housing market and that they consist of the city’s main housing demand; while rural migrants are part of the informal housing sector, which is not directly characterized here, although they might exert some pressure on the urban land supply for formal residential housing development.\(^{10}\)

Other variables used in this section, such as city’s historical average wages for staff and workers, province’s historical employment size and population size, utilized foreign direct investment (FDI), etc., are either from National Urban Yearbook 1991 or National Statistical Yearbooks (1991 and 1996). The changes in other local economic variables during our study period are from National Urban Yearbooks 2003 and 2009. Pairwise distances between cities are calculated by ArcGIS. The 1990 and 2000 censuses are also used to construct the instrumental variables.

Finally, we deflate all the nominal variables in our regressions, using the yearly provincial CPI from National Statistical Yearbooks. Summary statistics are in Table 3.1.

The long-difference specification for estimating the population elasticity \(\phi_2\) comes from a simplified version of (6).

\[
(13) \Delta \ln p_{ij/020k} = \Delta_{020k} + \phi_2 \Delta \ln N_{ij/020k} + \epsilon_{ij},
\]

\(^{10}\) The rural migrants discussed here are actually referred to residents without local urban hukou who are poorly educated, poorly skilled, and poorly paid. This group of residents represents the majority of “floating population” in China, whose living conditions and working conditions are deeply affected by the barriers from hukou system. They are de facto excluded from urban formal housing market. Furthermore, there may be buyers from other cities for investment purposes. In this paper, we do not consider investment demand.
where $\Delta \ln P_{j0208}$ represents the housing price appreciation in city $j$ between 2002 and 2008, and $\Delta \ln N_{j0208}$ is the population growth among local urban hukou holders between 2002 and 2008. In (13), we include only the population growth. The other three terms in (6), in addition to $\varepsilon_j$, are thus all left in the error term $\varepsilon_{x_j}$ of (13). $\phi_2$ is expected to be positive.

$\varepsilon_{x_j}$ correlates with $\Delta \ln N_{j0208}$ due to several reasons. First, high housing prices might discourage potential permanent migrants from settling down in this city, which would lead to a downward bias in the OLS estimate (reverse causality). Second, the unobserved economics shocks that affect housing prices in the 2000s may be correlated with population growth as well (omitted variables). Third, the changes in the other fundamentals could also correlate with $\Delta \ln N_{j0208}$. For example, the city government may accommodate a positive population growth of local urban hukou holders with an increase in urban land supply, thereby leading to a downward bias in the OLS estimate. Column 1 of Table 3.3 reports the OLS estimate for specification (13), and the coefficient of population growth is insignificant, reflecting a downward bias resulting from the endogeneity issues discussed above: the population growth of local urban hukou holders in the 2000s could be negatively affected by the rise in local housing costs, and it also may force local city government to raise urban land supply. When controlling for the changes of the other three fundamental factors in column 2, this coefficient drops to be insignificantly negative.
We use historical and geographic IVs to deal with these endogeneity issues. The ideal instruments should have strong power in predicting variation in $\Delta \ln N_{j2008}$, but be exclusive to the observed and unobserved contemporary shocks in the error term. To find such IVs, we first explore the sources for the population growth of local urban *hukou* holders between 2002 and 2008. There are three possible sources: 1) *hukou* migrants from the same province, 2) *hukou* migrants from other provinces, and 3) the natural growth among local urban *hukou* population. Accordingly, we specify the first stage regression as below:

$$
\begin{align*}
\Delta \ln N_{j2008} &= \theta_0 + \theta_1 \ln N_{p,j1990} + \theta_2 M_{j2000} + \theta_3 \text{shra}_{j1990} + \theta_4 \ln N_{j2002} + \nu_j.
\end{align*}
$$

$\ln N_{p,j1990}$ is the 1990 population size of the province in which city $j$ is located, which relates to the population growth of local urban *hukou* holders from within-province *hukou* migration in the study period (source 1 above). Either economic shocks in the 2000s or the rise in local housing prices is unlikely to affect the historical population distribution.

In addition, variation in the size of cross-province non-*hukou* (unregistered) migrant labor in city $j$ in 2000, denoted by $M_{j2000}$, could potentially capture the population growth of local urban *hukou* holders from cross-province migration (source 2) during 2002-2008. (Appendix C presents details on the definition of the cross-province non-*hukou* migration.) The larger the number of non-*hukou* migrants piling up in 2000, the more of them became legalized after 2000, when the central government started to decentralize the *hukou* administration to local authorities. The issue of using the actual migrant stock as the instrumental variable is as follows. Non-*hukou* migrants were attracted to
destinations by economic opportunities in both the manufacturing and service sectors, and the demand-pull component of this migrant flow is likely to be correlated with economic shocks in destination cities in the 2000s. Therefore, instead of directly using $M_{j,2000}$, we exploit the geographic distribution of employment in potential sending provinces as well as the economic conditions in urban areas before thorough reforms on state-owned-enterprise (SOE) were launched in the 1990s (Au & Henderson, 2006), to predict non-hukou migrant flows from each potential sending province to destination city $j$. We assume that these geographic and pre-reform economic factors affect current housing price appreciation only through the growth in local urban hukou population after 2000. We then constructed the predicted cross-province non-hukou migrant labor size in 2000 in destination city $j$, represented by $\hat{M}_{j,2000}$, by aggregating its predicted migrant flows from all the 29 potential sending provinces. Details of constructing $\hat{M}_{j,2000}$ are in Appendix D.

$s_{shrfa,j,1990}$ is the share of women aged 25-34 among those aged 25-44 in city $j$ in 1990 from the 1990 census. Relative to cohort 35-44, the fertility behaviors of cohort 25-34 were largely affected (in a negative way) by the One-Child policy introduced in 1979 (Short, Zhai, 1998). Their children born around early 1980s turned to child-bearing age in early 2000s. Therefore, a city with a relatively high $s_{shrfa,j,1990}$ is likely to have a relatively low natural population growth (source 3) in the 2000s. We assume that variation in the female age structure across cities in 1990 is exogenous to the error term in (13). In
addition, in (14) the initial local urban *hukou* population size in 2002, \( \ln N_{j2002} \), is meant to pick up the negative effect of initial population size on the growth rate.

Column 3 of Table 3.3 presents the first-stage results. As expected, the coefficients of \( \ln N_{j1990} \) and \( \hat{M}_{j2000} \) are significantly positive, and the coefficient of \( shrfa_{j1990} \) is significantly negative. The F-statistic is greater than ten, indicating that variations in our four instrumental variables can sufficiently capture variation in the population growth among local urban *hukou* holders during the 2002-2008 time period. Compared with the insignificant OLS estimate, the 2SLS estimate is significantly positive, with a point estimate of about 0.542 as shown in Column 4 of Table 3.3.

However, the constructed *hukou* population growth, \( \Delta \ln \hat{N}_{j2008} \), could still be correlated with changes in the other three fundamentals (i.e., income, land supply and construction cost) as well as the unobserved local shocks that are left in the error term of (13). Table 3.4 describes the correlations between the constructed urban *hukou* population growth from the first stage (\( \Delta \ln \hat{N}_{j2008} \)) and the other local shocks, including the other demand and supply shifters as well as seven observed local economic variables (utilized FDI, gross domestic product, gross industrial output value of enterprises above a designated size, sale revenue of product, value-added tax revenue, investment in fixed asset and the number of firms above a designated size). The results show that the positive shifters for housing price in equilibrium (i.e., income and construction cost) both negatively correlate with \( \Delta \ln \hat{N}_{j2008} \), while the negative supply shifter (i.e., land supply) is positively (but insignificantly) correlated with \( \Delta \ln \hat{N}_{j2008} \), which implies that \( \Delta \ln \hat{N}_{j2008} \) is negatively
correlated with \((\phi_1 \Delta \ln w_{j0208} - \phi_3 \Delta \ln L_{j0208} + \phi_4 \Delta \ln P_{Mj0208})\). The correlations between 
\(\Delta \ln \hat{N}_{j0208}\) and the other local economic variables are generally insignificant, except for the utilized FDI. The change in utilized FDI is more likely to be a positive housing price shifter. Therefore, we consider our point estimate in column 4 of Table 3.3 as a lower-bound estimate for population elasticity \(\phi_2\), and our calculated equilibrium housing price appreciation in Section 3.5 is a conservative estimate for the impact of fundamentals.

### 3.6 Results and Discussions

How much of the housing price appreciation can be explained by changes in fundamentals? In which city has housing price appreciation significantly deviated from changes in fundamentals? To answer these questions, in this section, we calculate the equilibrium housing price appreciation rates from the changes in fundamental factors of demand and supply between 2002 and 2008, based on the equilibrium relationship between housing price appreciation and the fundamentals presented in (5).

Data sources for the fundamental variables used in our calculations are as follows. The total area of land whose leaseholds were sold to developers by the city government between 2003 and 2007 measures the new increase in city’s urban land stock, and the total built-up land area within the city proper (i.e., land area under complete construction) in 2002 and 2003 measures the initial urban land stocks. The numbers are from National Land Resource Yearbooks and National Urban Yearbooks respectively. The growth rate of construction cost is constructed by the provincial price indices of investment in fixed assets for newly completed buildings and the provincial average labor cost in
construction sector, weighted by capital and labor shares, which are 0.8 and 0.2 in China (Jin, Liu, and Li, 2005) respectively, and deflated by the provincial CPI. Data are from National Statistical Yearbooks (2003-2009). Wage income is from National Urban Yearbooks. Hukou population is also reported by National Urban Yearbooks (Table 3.1).

We recover five coefficients in (5): income elasticity $\phi_1$, population elasticity $\phi_2$, land supply elasticities $\phi_3$, construction cost elasticity $\phi_4$, and the common growth factor $\Delta\phi_0$. We directly obtain the lower bound of $\phi_2$ in Section 3.5. $\phi_1$, $\phi_3$ and $\phi_4$ are inferred from the estimated land-share parameter in the housing production function $(1-\alpha)$ as well as two demand side elasticities (i.e., income elasticity $\delta_1$ and price elasticity $\delta_3$ in (3)) borrowed from the literature according to (4). The common growth factor $\Delta\phi_0$ in (5) is calculated by the difference between the average of the actual appreciation rates and the average of the calculated appreciation rates across the 35 cities.

Polinsky and Ellwood (1979) find that the permanent income elasticity of housing demand is around 0.8, Harmon (1989) places it at one for the long-run housing consumption, Epple and Sieg (1999) estimate it at 0.938, and Epple, Peress, and Sieg (2010) suggest that it is 0.787. For the price elasticity, it ranges from -0.5 to -0.9 (e.g., Polinsky and Ellwood, 1979; Epple and Sieg, 1999; Epple, Peress, Sieg, 2010). In the case of China, Chow and Niu (2010) find that the long-run income elasticity of demand for urban housing to be about 1.0, and the price elasticity of demand to be between -0.5
and -0.6. We place the income elasticity of demand $\delta_1$ at 1.0 and the price elasticity of demand $\delta_3$ at 0.5.\(^{11}\)

Thus, for $\phi_1$, the estimate is \( (1 - \hat{\alpha}) / (\hat{\alpha} + 0.5(1 - \hat{\alpha})) \), applying the estimate of the land-share parameter \((1 - \alpha)\) from section 3.4 and the values of $\delta_1$ and $\delta_3$ borrowed from the literature. The estimates of $\phi_3$ and $\phi_4$ are \( (1 - \hat{\alpha}) / (\hat{\alpha} + 0.5(1 - \hat{\alpha})) \) and \( \hat{\alpha} / (\hat{\alpha} + 0.5(1 - \hat{\alpha})) \), respectively. Applying Delta Method, we can derive their asymptotic distributions (Casella and Berger, 2002). We then generate the upper bounds of the 95-percent confidence intervals for the calculated housing price appreciation rates (see Appendix E for details).

Figure 3.2 plots the 95-percent confidence bands of the calculated housing price appreciation between 2002 and 2008 (two dash lines), ordered by city-size rankings (measured by the population of local urban hukou holders in 2002). Solid dots are actual appreciation rates during this period located within the 95-percent confidence intervals; solid diamonds represent actual appreciation rates above their upper bounds; crosses are actual appreciation rates below their lower bounds. For 18 (type-A cities) out of the 35 cities in our sample, the calculated housing price growth rates fall into the ranges of their 95-percent confidence intervals. Another ten cities (type-B cities) are below their lower bounds. There are seven cities that are above their upper bounds (type-C cities): Beijing, Fuzhou, Hangzhou, Ningbo, Tianjin, Wuhan, and Xiamen. Among these seven cities, Fuzhou, Ningbo, and Xiamen have their actual appreciation rates significantly higher.

\(^{11}\) We experiment with different values for these two parameters as robustness check. Specifically, we let $\delta_1 \in \{0.8, 0.9, 1.05, 1.1\}$ and $\delta_3 \in \{0.4, 0.45, 0.5, 0.55, 0.6\}$. The calculation results are all similar.
than their upper bounds (more than 120-percent higher); while, the actual appreciation rates in the other four cities are moderately higher (around 10 to 20-percent higher).

Type-C cities seem to have “overly high” appreciation rates, which suggest a risk for “bubbles.” Our identification of the seven Type-C cities is consistent with the study of Wu, Gyourko, and Deng (2011) which shows that housing markets in several big Chinese cities are very risky. In Type-A and Type-B cities, changes in fundamentals can largely explain actual housing price appreciation.

What is happening in Type-C cities? Cai, Cao and Zhang (2012) examine the investment demand for residential housing in urban China, using household survey data. They find that higher expected capital gains encourage investment demand in housing markets from rich urban households, in the form of purchasing a more valuable primary residence and/or a second house when facing a higher housing price growth rate. We find that, in these seven cities, housing price growth rates are higher than the other major cities during the first half of the study period, which might help form a higher expected capital gain from owing housing as assets. Moreover, households in these seven cities might be relatively rich and therefore more likely to have investment demand for housing (five of them have city average income above the 75 percentile of the initial income distribution). Therefore, we may expect the investment demand is another important driving factor for housing price appreciation in these seven cities, which is not characterized in our model. A thorough examination of potential risks faced by Type-C cities is beyond the scope of this paper and we leave it to future works.
Another point worth noting is that the common growth rate takes up around 40-percent of the actual appreciation rate on average across 35 cities. This common growth rate might be brought about by the institutional change due to the reform in Chinese urban land markets between 2002 and 2008. Since 1988, use rights for vacant land in the city area were allocated through either land leasing, which was done by public auction, tendering and negotiation, or “land development rights (LDR).” The latter had allowed large traditional holders (e.g., state-owned enterprises) to privately sell off their own land use rights in a secondary market, which, in fact, facilitated the co-operation between local government and local state-owned enterprises/developers in a fashion of reciprocity when redeveloping the “brown-field” in the city center (Zhu, 2004, 2005). In the 1990’s, most use-right allocations were done either by “negotiation” in a hidden process or in the form of “LDR.” A 2002 law banned negotiated sales for the profitable development project by land bureaus, with the last date for any negotiated sales being August 31, 2004. Another reform in 2002 banned the secondary market for “LDR.” Since 2004, all urban leasehold sales for purely private development are done through public auctions. The government has completely monopolized the land supply; meanwhile, available urban land for development gets scarcer. These potentially aggravate land developers’ expectations of the decline in urban land supply in the future, which, to some extent, has been reflected by the soaring land prices since 2004. It is believed that many land developers speculatively hold back the purchased land from development, which, in effect, restricts the residential housing supply and pushes up the housing prices across all cities.
3.7 Conclusions

In this paper, we investigate the role of fundamental factors in China’s urban housing markets. Our results suggest that for most of the cities in our sample, actual housing price appreciation can be largely explained by changes in fundamentals, such as land supply, urban hukou population, wage income, and construction cost. Seven cities seem to have “overly high” actual appreciation rates, which indicate a higher risk of bubbles in these cities. One caveat of this paper is that while we can tell whether people in some cities are overpaying for their houses relative to other cities, we cannot conclude whether national housing prices are too high. To evaluate whether housing price levels are reasonable, we need to apply a more financial approach that relies on there being no predictable excess return to being an owner relative to being a renter (“price-to-rent ratio”), which is beyond the scope of this paper.
### 3.8 Tables

<table>
<thead>
<tr>
<th>Variable label</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{v}$</td>
<td>Housing value per unit of land, yuan per sqm.</td>
<td>20191.997</td>
<td>21510.844</td>
</tr>
<tr>
<td>$\tilde{p}_l$</td>
<td>Price per unit of land, yuan per sqm.</td>
<td>4464.492</td>
<td>5254.265</td>
</tr>
<tr>
<td>dist</td>
<td>Distance to CBD, km</td>
<td>11.376</td>
<td>11.606</td>
</tr>
<tr>
<td>railway</td>
<td>Whether there exists a railway (subway/light rail) within 2.5-km radius.</td>
<td>0.455</td>
<td>0.499</td>
</tr>
<tr>
<td>highway</td>
<td>Whether there exists a highway within 2.5-km radius.</td>
<td>0.576</td>
<td>0.495</td>
</tr>
<tr>
<td>auctiontype</td>
<td>Whether the land sale is done through Two-Stage auction.</td>
<td>0.774</td>
<td>0.418</td>
</tr>
</tbody>
</table>

#### Residential development level characteristics

<table>
<thead>
<tr>
<th>Variable label</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{M}_{2000}$</td>
<td>The predicted cross-province migrant labor size in 2000. Use estimates from Table 3.5</td>
<td>4.3264</td>
<td>7.4913</td>
</tr>
<tr>
<td>$\ln N_{p1990}$</td>
<td>The logarithm of 1990 population size in the province of destination (10000 persons).</td>
<td>8.1137</td>
<td>0.78</td>
</tr>
<tr>
<td>$\ln N_{2002}$</td>
<td>The logarithm of city's legal local urban citizen population size in 2002 (10000 persons).</td>
<td>5.46</td>
<td>0.693</td>
</tr>
<tr>
<td>shrfa$_{1990}$</td>
<td>The share of female aged 25-34 among ones aged 25-44 in 1990 calculated from the 1990 census.</td>
<td>0.5723</td>
<td>0.0319</td>
</tr>
<tr>
<td>$\Delta \ln p_{2008}$</td>
<td>The percent change in price of residential housing sales, 2002-2008.</td>
<td>0.5533</td>
<td>0.2072</td>
</tr>
<tr>
<td>$\Delta \ln w_{2008}$</td>
<td>The percent change in city average wage, 2002-2008.</td>
<td>0.615</td>
<td>0.1462</td>
</tr>
<tr>
<td>$\Delta \ln N_{2008}$</td>
<td>The percent change in local urban hukou holders, 2002-2008.</td>
<td>0.2541</td>
<td>0.1523</td>
</tr>
<tr>
<td>$\Delta \ln L_{2008}$</td>
<td>The percent change in urban land stock, 2002-2008.</td>
<td>0.422</td>
<td>0.232</td>
</tr>
<tr>
<td>$\Delta \ln p_{M2008}$</td>
<td>The percent change in construction cost between 2002 and 2008.</td>
<td>0.2256</td>
<td>0.0757</td>
</tr>
</tbody>
</table>
Table 3.2: Housing Production Function Estimation

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>$\tilde{v}$</th>
<th>$\tilde{P}_i$</th>
<th>$\tilde{P}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variable</td>
<td>FS</td>
<td>OLS</td>
<td>2SLS</td>
</tr>
<tr>
<td>$\tilde{v}$</td>
<td>***0.287</td>
<td>***0.323</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>$\tilde{v} ,* \text{auctiontype}$</td>
<td>***-0.119</td>
<td>***-0.118</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>$\text{dist}$</td>
<td>***-348.562</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(77.047)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{railway}$</td>
<td>**4388.828</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1899.987)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{highway}$</td>
<td>1647.277</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1912.329)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{auctiontype}$</td>
<td>**3740.127</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1539.120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>F-stat for weak IV</td>
<td>34.812</td>
<td>34.812</td>
<td></td>
</tr>
<tr>
<td>Hansen J-stat</td>
<td>3.630</td>
<td>3.630</td>
<td></td>
</tr>
<tr>
<td>p-value of Hansen J</td>
<td>0.163</td>
<td>0.163</td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>505</td>
<td>505</td>
<td>505</td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. The table reports estimates of the equilibrium locus which characterizes the relationship between the price of land and the value of housing per unit of land.
Table 3.3: Impact of Population Growth on Housing Price Appreciation

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>$\Delta \ln p_{\text{a0208}}$</th>
<th>$\Delta \ln p_{\text{a0208}}$</th>
<th>$\Delta \ln N_{\text{a0208}}$</th>
<th>$\Delta \ln p_{\text{a0208}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 OLS</td>
<td>2 OLS</td>
<td>3 FS</td>
<td>4 2SLS</td>
</tr>
<tr>
<td>$\Delta \ln N_{\text{a0208}}$</td>
<td>0.187 (0.190)</td>
<td>-0.052 (0.301)</td>
<td>**0.542 (0.252)</td>
<td></td>
</tr>
<tr>
<td>$\Delta \ln w_{\text{a0208}}$</td>
<td>-0.546 (0.452)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \ln \bar{L}_{\text{a0208}}$</td>
<td>0.201 (0.134)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta \ln p_{\text{M}_{\text{a0208}}}$</td>
<td>0.763 (0.954)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ln N_{p_{\text{a1990}}}$</td>
<td>*0.059 (0.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{M}_{\text{a2000}}$</td>
<td>***0.013 (0.002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shrfa_{a1990}</td>
<td>*-1.645 (0.839)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\ln N_{p_{\text{a2002}}}$</td>
<td>***-0.138 (0.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansen J-stat</td>
<td>3.054 3.054</td>
<td>3.054 3.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value of Hansen J</td>
<td>0.383 0.383</td>
<td>0.383 0.383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>35 35</td>
<td>35 35</td>
<td>35 35</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses.
<table>
<thead>
<tr>
<th>Dep. var.</th>
<th>$\Delta \ln w$</th>
<th>$\Delta \ln L$</th>
<th>$\Delta \ln p_M$</th>
<th>$\Delta \ln fdi$</th>
<th>$\Delta \ln gdp$</th>
<th>$\Delta \ln giov$</th>
<th>$\Delta \ln salerp$</th>
<th>$\Delta \ln vatax$</th>
<th>$\Delta \ln invfa$</th>
<th>$\Delta \ln numfirm$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \ln N_{0208}$</td>
<td>***-0.819</td>
<td>0.38</td>
<td>**-0.233</td>
<td>-1.816*</td>
<td>0.209</td>
<td>-0.265</td>
<td>-0.246</td>
<td>-0.343</td>
<td>-0.359</td>
<td>0.652</td>
</tr>
<tr>
<td>(0.151)</td>
<td>(0.584)</td>
<td>(0.097)</td>
<td>(0.947)</td>
<td>(0.211)</td>
<td>(0.416)</td>
<td>(0.357)</td>
<td>(0.491)</td>
<td>(0.948)</td>
<td>(0.585)</td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses. Standard errors are adjusted for the sequential estimation. The time span is between 2002 and 2008. $\Delta \ln w$ is the wage growth rate. $\Delta \ln L$ measures the growth rate of urban land supply for residential use. $\Delta \ln p_M$ represents the growth rate of construction cost. $\Delta \ln fdi$ is the growth rate of utilized FDI. $\Delta \ln gdp$ is the growth rate in gross domestic product. $\Delta \ln giov$ represents the growth rate in gross industrial output value of enterprises above a designated size. $\Delta \ln salerp$ is the growth rate in the sale revenue of product. $\Delta \ln vatax$ is the growth rate in value-added tax revenue. $\Delta \ln invfa$ is the growth rate in investment in fixed asset. $\Delta \ln numfirm$ represents the growth rate in the number of firms above a designated size. $\Delta \ln N_{0208}$ is the constructed population growth rate between 2002 and 2008 from the first stage, in which the instruments are $\ln N_{p,1990}$, $M_{2000}$, $\ln N_{2002}$, and $shrfa_{1990}$. 
3.9 Figures

Figure 3.1: Trends in Cities’ Average Sale Prices of Residential Housing

Notes: Housing prices are deflated to currency in 2005 by provincial CPI indices from National Statistical Yearbooks. The cities at or above the 90th percentile in the distribution of housing price appreciation between 2002 and 2008 are Beijing, Fuzhou, Ningbo and Xiamen. The cities at or below the 10th percentile in the distribution of housing price appreciation between 2002 and 2008 are Ha’erbin, Kunming, Shenyang and Yinchuan.
Figure 3.2: Calculated and Actual Housing Price Appreciation between 2002 and 2008
3.10 Appendix A: Data for housing production function estimation

The land auction data set consists of 13,030 land-parcel sales for 2001-2009 across 60 cities, among which 6,294 are for residential use. Observations are properties put up for auction with completed transactions later. The *loupan* data set covers 13,083 development projects in April 2009 across 55 cities. It also has history information for part of these developments.

In the raw data of matched land parcel sales and *loupan*, we have 797 matched pairs between land parcel sales and *loupan* developments. We drop projects with different house types (villa versus apartment) but the same floor-area-ratio. We treat multiple observations as one project if they are using identical land parcel information and keep the observation with the lowest price per square meter of floor space. We treat multiple observations with same project name and different land parcel information as separate projects. We also drop observations with either missing or unreasonable floor-area-ratio. We drop observations with land parcels transacted through sealed-bid auctions. Finally, we drop observations with missing information on prices and the size of land parcels. We end up with 505 observations for the final regression.
3.11 Appendix B: Trends of Non-agricultural *hukou* Population for the 35 Cities and Big Events since 1949

Figure 3.3 shows the trend of the non-agricultural *hukou* population for these 35 cities. Before the national industrial policy was fully phased in (1949-1957), spatial mobility was allowed and there was an increase in the nonagricultural population in this period. In 1958, the *hukou* system was designed to curtail urbanization (to ensure enough labor supply in the rural sector to support the raw-material needs in heavy industry in urban areas); in this period, China had a high industrial growth rate but a low growth rate of urbanization (even negative). Economic reform has been phased in since 1978 in rural areas, and there has been a big increase in the nonagricultural population—largely due to the increase in productivity in the agricultural sector—and, therefore, a labor shift from farm work to manufacturing in TVE (Town-village enterprises) in cities and towns nearby. Massive economic reform in urban areas started in the early 1990s, and since then, the growth rate of the nonagricultural population has accelerated. It has become even higher since 2000, when the state started to decentralize the *hukou* decision to local authorities. The population of local urban *hukou* holders has increased dramatically since 2000.
Figure 3.3: Trends of Non-agricultural hukou Population for the 35 Cities and Big Events since 1949

Notes: The dash line represents the average population size of local urban hukou holders in each year of cities at and above the 90th percentile in the distribution of 2008 city size distribution (Beijing, Chongqing, Guangzhou and Shanghai). The dot line represents the average population size of local urban hukou holders in each year of cities at and below the 10th percentile in the distribution of 2008 city size distribution (Haikou, Huhehaote, Xining and Yinchuan). The solid line represents the average city size in each year of the 35 cities.
3.12 Appendix C: Definition of cross-province non-\textit{hukou} migrant workers and the 2000 census

The cross-province non-\textit{hukou} migrant is the person whose \textit{hukou} residence is different from his current residence and is located in another province. According to the official statistics, non-\textit{hukou} migration accounts for the major proportion of the total internal migration.

Labor force is defined as individuals who 1) are aged 15-65; 2) are not students in 2000; 3) either have a job between October 25 and October 31, 2000, or are on sabbatical, in training, on seasonal rest, or are searching for jobs.

Micro-level data of 0.1\% sample of the China 2000 Census are used to calculate the historical cross-province non-\textit{hukou} migrant labor-force size for 35 destination cities. Based on the actual provinces of origin, we generate a common set of provinces of origin, excluding the province of destination; that is, for each destination city, there are 29 provinces of origin. The province of origin is defined as the location of \textit{hukou}. The city of origin cannot be identified because of the limited information in the 2000 census. Therefore, we assume that people compared the potential city of destination and the capital city in their province of origin when making cross-province migration decisions.

We weight the original number of historical cross-province non-\textit{hukou} migrant workers in each destination from a specific province of origin by the provincial-level sampling weights for the 2000 census.
3.13 Appendix D: Prediction of cross-province non-\textit{hukou} migrant workers in 2000

We estimate migrant flow between each potential sending province \( p \) and destination city \( j \) as a function of geographic employment distribution and the pre-reform economic conditions, based on the Gravity model (Anderson and van Wincoop, 2003) in trade theory:

\[
M_{jp} = \lambda_0 + \lambda_1 \ln(\text{dist}_j) + \lambda_2 \ln(\text{emp}_{p1995}) \times \ln(\text{dist}_j) + \lambda_3 \ln(FDI_{j1990}) + \lambda_4 \ln(FDI_{j1990}) \times \ln(\text{dist}_j) + \lambda_5 \text{wagediff}_{j1990} \\
+ \lambda_6 \text{wagediff}_{j1990} \times \ln(\text{dist}_j) + \sum_{k=1}^{3} \lambda_k \text{SEZ}_j^k + \delta_p + \omega_{jp}
\]

\( j = 1, \ldots, 35 \). \( p \) is drawn from a set of 29 potential sending provinces that excludes the province in which \( j \) is located in. \( M_{jp} \) is the non-\textit{hukou} migrant labor flow from sending province \( p \) to destination city \( j \) measured from the 2000 census and adjusted by provincial sampling weights. Geographic population distribution is captured by \( \ln(\text{emp}_{p1995}) \times \ln(\text{dist}_j) \), which is the logarithm of the employment size in potential sending provinces in 1995, interacted with the logarithm of distance between the destination city \( j \) and the capital city in sending province. The economic conditions before SOE reforms are characterized by \( \ln(FDI_{j1990}) \) and \( \text{wagediff}_{j1990} \), which are the logarithm of utilized FDI in destination \( j \) in 1990 and the difference in logarithms of wages in the non-state and non-collective sectors in 1990 between destination city \( j \) and capital cities in sending provinces, respectively. We include their interaction terms with
distances to capture the effects of geographic proximity on the “pull factors.” We also include the special economic zone (SEZ) dummies, allowing for the heterogeneous effects of being SEZs. Moreover, we control for the origin fixed effects $\delta_p$, as some provinces, such as Sichuan, Henan, and Anhui, are historically the main origins of migrants. Table 3.5 reports the results. As expected, distance weakens the effect of employment size of sending province. Both utilized FDI and wage differential in 1990 work as positive “pull factors,” while distance weakens these pulling effects. The main effect of distance is significantly positive, which suggests that pair distance might capture some unobserved pair effect, such as an historical migration route. Based on the estimates of coefficients in (A1), we construct the cross-province non-hukou migrant labor size in city $j$ from province $p$, $\hat{M}_{jp}$. We then aggregate this predicted migrant flow over all the 29 potential sending provinces to obtain the predicted cross-province non-hukou migrant labor size in city $j$, $\hat{M}_{j2000}$, which is used as one of the IVs in (14).
Table 3.5: Historical non-hukou Migrant Flows

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{jp2000} )</td>
<td>( \ln\text{emp}<em>{p1995} ) * ( \ln\text{dist}</em>{jp} )</td>
<td><strong>-0.0996</strong></td>
<td>(0.0382)</td>
</tr>
<tr>
<td></td>
<td>( \ln\text{dist}_{jp} )</td>
<td><strong>0.9896</strong></td>
<td>(0.3686)</td>
</tr>
<tr>
<td></td>
<td>( \ln\text{FDI}_{j1990} )</td>
<td><strong>0.9874</strong></td>
<td>(0.3688)</td>
</tr>
<tr>
<td></td>
<td>( \ln\text{FDI}<em>{j1990} ) * ( \ln\text{dist}</em>{jp} )</td>
<td><strong>-0.0699</strong></td>
<td>(0.0262)</td>
</tr>
<tr>
<td></td>
<td>( \text{wagediff}_{jp1990} )</td>
<td>*<strong>10.0319</strong></td>
<td>(2.4862)</td>
</tr>
<tr>
<td></td>
<td>( \text{wagediff}<em>{jp1990} ) * ( \ln\text{dist}</em>{jp} )</td>
<td>*<strong>-0.6809</strong></td>
<td>(0.1743)</td>
</tr>
<tr>
<td></td>
<td>( \text{SEZ}_{Shenzhen} )</td>
<td>*<strong>1.0245</strong></td>
<td>(0.3199)</td>
</tr>
<tr>
<td></td>
<td>( \text{SEZ}_{Xiamen} )</td>
<td><strong>-0.1667</strong></td>
<td>(0.0648)</td>
</tr>
<tr>
<td></td>
<td>( \text{SEZ}_{Haikou} )</td>
<td>-0.1082</td>
<td>(0.0741)</td>
</tr>
<tr>
<td></td>
<td>Origin fixed effects</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R-squared</td>
<td>0.3583</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>1015</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *significance at 10%; **significance at 5%; ***significance at 1%. Heteroskedasticity-robust standard errors are in parentheses, and they are clustered by the sending province. \( \ln\text{emp}_{p1995} \) is the logarithm of the employment size in potential sending province in 1995. \( \ln\text{dist}_{jp} \) is the logarithm of distance between destination city and capital city in sending province. \( \ln\text{FDI}_{j1990} \) represents the logarithm of utilized FDI in destination city in 1990. \( \text{wagediff}_{jp1990} \) is the difference in logarithms of wages in the non-state and non-collective sectors in 1990 between destination city and capital city in sending province. \( \text{SEZ}_{Shenzhen} \), \( \text{SEZ}_{Xiamen} \), and \( \text{SEZ}_{Haikou} \) are SEZ dummies. \( M_{jp2000} \) is the weighted cross-province migrant labor flow in 2000, from province \( p \) to destination city \( j \).
3.14 Appendix E: Generate the 95-percent confidence intervals for the calculated housing price appreciation rates

The mean of the calculated housing price appreciation between 2002 and 2008 for city $j$ is

$$\overline{\Delta \ln \hat{p}_{hj}} = \Delta \hat{\phi}_0 + \hat{\phi}_1 \Delta \ln w_j + \hat{\phi}_2 \Delta \ln N_j - \hat{\phi}_4 \Delta \ln L_j + \hat{\phi}_4 \Delta \ln p_M$$

The variance of the calculated housing price appreciation is

$$sd_j^2 = (\Delta \ln w_j - \Delta \ln \overline{L_j})^2 \text{var}(\hat{\phi}_1) + (\Delta \ln p_M)^2 \text{var}(\hat{\phi}_4)$$

$$+ 2 \text{cov}(\hat{\phi}_1, \hat{\phi}_4)(\Delta \ln w_j - \Delta \ln \overline{L_j})\Delta \ln p_M + (\Delta \ln N_j)^2 \text{var}(\hat{\phi}_4).$$

The 95-percent confidence interval for city $j$ is given by:

$$[\Delta \ln \hat{p}_{hj} - 1.96 \times sd_j, \Delta \ln \hat{p}_{hj} + 1.96 \times sd_j].$$
Bibliography


De la Roca, Jorge, Gianmarco Ottaviano, and Diego Puga. 2012. “City of Dreams: Self-Confidence, Ability, Luck, and the Experience and Opportunities Offered by Cities.”


