Essays on the Labor Market, Housing and Children's Health

by

Xi Yang

A dissertation submitted to The Johns Hopkins University in conformity

with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

August, 2014

Abstract

This dissertation studies the consequences of housing and migration decisions from three perspectives using data from developed and developing countries. Chapter 2 studies the effects of home ownership on unemployment in U.S. by developing and estimating a dynamic job search model. The estimation results show that home ownership decreases unemployment duration by about nine days. Young workers with low education levels are most likely affected—their unemployment durations are shortened by about 14 days. Counterfactual experiments show that eliminating the mortgage interest deduction is not an efficient policy for reducing unemployment and actually increases unemployment duration by three days.

Chapter 3 studies the impact of mortgage status on the female labor supply in U.S.. By looking at married women in SIPP 1996 panel data from 1996 to 2000, I find a positive and significant effect of home mortgage on the female labor supply. A large mortgage increases not only female's labor participate rate, but also their hours of work. Subsample results show that women with limited household wealth are more likely to be affected by the home mortgage.

Chapter 4 studies the impact of labor migration on children's health in China. We use China Health and Nutrition Survey (CHNS) in 2000, 2004, 2006, and 2009 to identify the impact of parents' migration on the health outcomes of children in rural China. The measurements of child health outcomes are weight-for-age Z-score (WAZ), height-for-age Z-score (HAZ), nutrient intake (consumption of calories and protein), the number of immunization shots that children get in the survey year and child-care. We found there were few significant effects of parents' migration on child health outcomes.

Keywords:	Job Search, Female Labor Supply, Home Ownership,
	Home Mortgage, Moving Costs, Endogeneity, Self-selection
	Children's health, Labor Migration, Fixed effects
JEL Classification:	C13 C23 I15 J22 J61 J64 R21
Advisors:	Professor Robert Moffitt
	Professor Yuya Sasaki

Acknowledgements

First, I am deeply indebted to my advisor, Professor Robert Moffitt, for his guidance and encouragement on this project, and for his support and understanding during my studies at Johns Hopkins University.

I thank Professor Yuya Sasaki, Professor Yingyao Hu, and Professor Hulya Eraslan for their guidance and support, to Professor Christopher D. Carroll for his valuable comments, to Dr. Kevin Thom for his suggestions on computation methods, to Dr. Guofang Huang, Dr. Haomiao Yu, Dr. Kai Liu, and Dr. Yonghong An for their comments and moral support.

I thank all graduates students at Hopkins, especially Gizem Kosar, Gwyn Pauley, Mikhail Smirnov, Alexandria Zhang, Ruli Xiao, Hou Wang and Xu Lu, who come to my presentation, help revising my drafts and be supportive.

I could not thank my dear parents and my parents-in-law more. They fly thousands of miles from China to Baltimore just to help me. I would not manage without their love and support.

Finally, I dedicate this dissertation to my husband, Yase Ge and my adorable little boy Reid who just turned eight months.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

How to live, rent or own, and where to live, urban or rural, are two of the most important decisions people make in their life. Those decisions largely shape people's living conditions and more importantly, have significant impacts on their labor market outcomes as well as their family member's welfare conditions. This dissertation studies the consequences of housing and migration decisions from three perspectives using data from both developed and developing countries.

Chapter 2 studies the effects of home ownership on unemployment in U.S. by developing and estimating a dynamic job search model. There is a growing concern among U.S. policymakers that home ownership may inhibit labor market matching because it restricts homeowners' ability to relocate for non-local job vacancies. This chapter argues that this concern is misplaced because it neglects the fact that homeowners are more likely to accept local job offers and be employed. The model in this chapter distinguishes between local and non-local job offers. The overall unemployment duration is estimated by summing the spells that end with the acceptance of local jobs and those end with non-local jobs. To address the self-selection problem of being a homeowner, this paper models home ownership and job search decisions jointly and adopts state-level mortgage interest deductions as an exclusion restriction in the estimation. The estimation results show that home ownership decreases unemployment duration by about nine days. Young workers with low education levels are most likely affected—their unemployment durations are shortened by about 14 days. Counterfactual experiments show that both the magnitude of moving costs and the labor market environment are important determining factors of the causal effects. Counterfactual experiments show that eliminating the mortgage interest deduction is not an efficient policy for reducing unemployment and actually increases unemployment duration by three days.

Chapter 3 studies the impact of mortgage status on the female labor supply in U.S.. To solve the self-selection problem that women with better labor market prospects are more likely to overcome financial constraints and commit to a larger home mortgage, I adopt the state-level mortgage interest deduction as an instrumental variable. By looking at married women in SIPP 1996 panel data from 1996 to 2000, I find a positive and significant effect of home mortgage on the female labor supply. A large mortgage increases not only female's labor participate rate of female, but also their hours of work. Subsample results show that women with limited household wealth are more likely to be affected by the home mortgage.

Chapter 4 studies the impact of labor migration on children's health in China. Labor migration, which frequently results in family separations, is widely known as one of the main ways of alleviating poverty in developing countries. In China, migrant workers helped build the Chinese dream in cities across the country. But for their children, who are left behind in the countryside, the potential health problems of their physical and social development is becoming a national issue. I use China Health and Nutrition Survey (CHNS) in 2000, 2004, 2006, and 2009 to identify the impact of parents' migration on the health outcomes of children in rural China. The measurements of child health outcomes are weight-for-age Z-score (WAZ), height-for-age Z-score (HAZ), nutrient intake (consumption of calories and protein), the number of immunization shots that children get in the survey year and child-care. To identify the effect of parental migration on child health, we instrumented parents' migration status with county level historical average migration rates. We found there were few significant effects of parents' migration on child health outcomes.

Chapter 2

The Effects of Home Ownership on Unemployment

2.1 Introduction

In the U.S., home ownership has long been viewed as part of the American Dream, and the tax code reflects this aspiration. Under the current U.S. income tax system, homeowners may deduct tax from mortgage interest, which is one of the largest federal tax expenditures. [1]Those policies are commonly justified by the positive externality of home ownership.[2]

The recent housing crisis, however, raises concerns[3] about home ownership by linking it with the nation's persistently high unemployment rate and long unemployment duration. These studies, largely propelled by Oswald (1996), Oswald

---

[1] The Adminstration's fiscal year 2014 budget, released in April 2013, estimated that the mortgage interest deductions would cost $93 billion in 2013 and $640 billion from 2014 to 2018.

[2]As mentioned in Glaeser and Shapiro (2003) and DiPasquale and Glaeser (1999), home ownership encourages investment in local amenities and social capital.

[3]See, e.g.,Ferreira, Gyourko, and Tracy (2010), Winkler (2010), Karahan and Rhee (2012), Valletta (2013) and Blanchflower and Oswald (2013).

(1997), argue that home ownership restricts homeowners' ability to relocate for non-local job vacancies, thus leading to longer unemployment durations. Convinced by this "mobility effect" argument, many policymakers believe that discouraging home ownership would be one efficient way to reduce labor market friction, therefore, reduce unemployment. However, policy implications in those studies are misleading because they neglect the responses of unemployed owners when they receive local job offers.

To complete the picture, this paper argues that homeowners have a stronger incentive than equally qualified renters to accept local job offers because moving costs lower their expected valuation of future job offers. This "incentive effect" leads to a higher transition rate into employment locally, even though the "mobility effect" lowers the transition rate into employment outside the local labor market. The net effects of home ownership on unemployment depend on the empirical magnitudes of each of these two effects. The literature emphasizes the effects of home ownership in the non-local market–the mobility story–while neglecting the effects in the local market–the incentive story. This paper incorporates both mechanisms in a job search model that allows agents to receive job offers from both non-local and local markets, and provides a more comprehensive evaluation of the effects of home ownership on unemployment.

To estimate the effects of home ownership on unemployment and to control for self-selection into home ownership, this paper jointly models and estimates the binary decision of owning and renting and the transitions between employment and unemployment. In addition, it includes state-level mortgage interest deductions as an exclusion restriction to add exogenous variation in home ownership status. The estimates show that, weighing the transition probability from employment to unemployment in both local and non-local markets, homeowners are more likely to leave unemployment but less likely to leave for a distant job offer. Overall, home ownership leads to a lower aggregate unemployment rate and shorter unemployment durations. That is, the concern that home ownership decreases labor market efficiency and keeps unemployment high is misplaced.

Compared with the standard job search model, two characteristics stand out in this model. First, instead of considering only the overall transition from unemployment to employment, this model distinguishes transitions from unemployment to two types of employment—employment in local jobs and employment in non-local jobs. These correspond to the "incentive effect" and the "mobility effect," respectively. The non-local offer is defined as a job opportunity that requires a residential move and triggers moving costs. An unemployed worker accepts a wage offer if and only if it brings higher utility than the expected utility of future offers. When unemployed, renters respond to local and non-local job offers in the same way because they are free to move. However, homeowners behave differently facing local and non-local job offers because moving costs alter their expectations of future offers. On the one hand, homeowners are more likely to accept local offers because the costs of moving reduce the attractiveness of future non-local offers and lead to lower expected values of future offers. On the other hand, they are more reluctant to accept non-local job offers unless these offers provide higher wages to compensate for the costs of a geographic move. The overall effects of home ownership on unemployment depend on the distribution of job offers across the two markets.

Second, the model allows the agent to decide whether to own or rent and specifies the correlation between unobserved preference for owning and unobserved skill endowments. This feature of the model provides an alternative strategy to address the endogeneity problem caused by reverse causality or self-selection. For example, employed people or people with higher ability are usually more likely to overcome financial constraints and become homeowners. Two approaches have been adopted in the literature to deal with this endogeneity problem. The first is the traditional instrumental variables method.[4] The second method exploits the occurrence of multiple unemployment spells.[5] Compared with these conventional methods, the structural approach provides an alternative identification strategy of the causal effect. Estimating structural parameters also makes it possible to directly test the mechanism behind this causal relationship in alternative environments for certain demographic groups and to conduct counterfactual policy analysis.[6]

Furthermore, this paper uses the state-level mortgage interest deduction as an exclusion restriction, which approximate the owning cost that exogenously affects the home ownership decision. This tax policy affects the home ownership decision but has no direct correlation with labor market transitions. This feature makes the identification not completely reliant on the functional form and distributional assumptions of the error terms, but also on the exclusion restriction.

The structural parameters of the theoretical model are recovered by the method of Maximum Simulated Likelihood using the Survey of Income and Program Participation (SIPP) 1996 panel. The structural approach requires estimates of a large number of parameters (31 in total). To avoid the potential problem of misspecification in the estimation, I conduct a robustness check by adopting the adaptive LASSO (least absolute shrinkage and selection operator) method. The model fits the data reasonably well. Estimating structural parameters allows me to simulate a group of renters, a control group for homeowners. Compared with them, unemployed own-

[4]See Van Leuvensteijn and Koning (2004) and Coulson and Fisher (2009).

[5]See Munch, Rosholm, and Svarer (2006).

[6]See Moffitt (2003), Heckman and Vytlacil (2007), Heckman and Urzua (2010) and Keane (2010) for more discussions of the IV approach.

ers are more likely to accept a local job offer and to reject a non-local offer. Overall, because a majority of job offers come from the local market, unemployed homeowners have a lower unemployment rate and shorter unemployment durations. By calculating this effect for different demographic groups, I find that young workers with low education levels are the most likely to be affected by home ownership. To investigate the important role of housing and labor market environments, home ownership and job search behaviors are simulated in two counterfactual experiments. In the first, home owners can move without cost. Simulations show that home ownership has little effect on unemployment. This leads to the conclusion that the causal effects are driven largely by the moving costs associated with home ownership. Second, I change the job offer distributions across local and non-local markets. In the case in which an unemployed agent can receive job offers only from the local labor market, home ownership has smaller effects on unemployment. In the case in which an unemployed agent can receive job offers only from the non-local labor market, homeowners are less likely to match with job vacancies, and home ownership indeed leads to higher unemployment, because this experiment eliminates the incentive mechanism. Lastly, I eliminate the mortgage interest deduction and find that it is not an efficient policy for reducing unemployment.

The paper is organized as follows. Section 2 summarizes the literature. Section 3 motivates the paper by describing the data sources and showing some empirical evidence. Section 4 presents the baseline life-cycle model with home ownership and employment dynamics. Section 5 provides some implications from the structural model. In Sections 6 and 7, I lay out the estimation strategy, the estimation results, as well as the fit and robustness of the model. Section 8 uses the model to generate counterfactual scenarios to assess the effects of home ownership on unemployment. Section 9 concludes.

2.2 Literature Review

In a series of papers, Oswald (1996),Oswald (1997) argues that home ownership is damaging to labor market outcomes using macro time series and cross-section data for OECD countries and regions within a number of those countries. His argument centers on immobility–homeowners are tied to their location and less likely to be matched with job vacancies when unemployed. At the aggregate level, this immobility translates into a positive correlation between an area's home ownership rate and its unemployment rate.

Inspired by this hypothesis, many researches attempted to test whether ownership has a negative impact on labor market outcomes. So far, however, no unanimous conclusion has been reached, either on the direction of the home ownership-unemployment correlation or on the the effectiveness of the immobility mechanism. Though some macro-data evidence provides some support for the hypothesis (Green and Hendershott (2001) and Blanchflower and Oswald (2013)), most micro-data results show that homeowners actually are less likely to be unemployed compared with renters(Coulson and Fisher (2002),Coulson and Fisher (2009),Flatau, Forbes, and Hendershott (2003),Munch, Rosholm, and Svarer (2006),Battu, Ma, and Phimister (2008), Van Leuvensteijn and Koning (2004)). The seemingly surprising fact based on this empirical evidence is that although home ownership does hamper geographical mobility substantially, it does not necessarily have negative effects on labor market outcomes. Recently, Oswald's hypothesis has been intensively cited by papers about the mortgage crisis. Those papers have led to heated debates on whether the mobility effects caused or exacerbated unemployment during the recent recession. For example, Winkler (2010), Ferreira, Gyourko, and Tracy (2010), and Karahan and Rhee (2012) are in favor of the mobility effect, while Schulhofer-Wohl (2011), Modestino and Dennett (2012) and Valletta (2013) argue that this negative effect on unemployment is limited.

Existing studies have two main limitations. First, though many empirical attempts have been made to study the relationship between home ownership and unemployment, few economic theories, besides Oswald's hypothesis, have been provided to explain this relationship. Although Munch, Rosholm, and Svarer (2006) mention the theoretical framework of the job search, they do not model the choice of home ownership, and their estimation is not based on the model. Head and Lloyd-Ellis (2012) develop a two-city model that allows for search frictions in both housing and labor markets, but they do not conduct any empirical work.

Second, the causal effects of home ownership on unemployment are hard to identify due to the endogeneity problem. The literature has adopted two approaches to address this endogeneity problem. The first is the traditional instrumental variable method (2SLS). For example, Van Leuvensteijn and Koning (2004) use the regional home ownership rate, and Coulson and Fisher (2009) use the state-level marginal tax rate as the instrumental variable. The second method identifies the causality by the existence of multiple unemployment spells (Munch, Rosholm, and Svarer (2006)). Though those reduced-form estimations are useful in quantifying the marginal treatment effect, little can be derived about the causal mechanism without proper theory, as mentioned in Heckman and Vytlacil (2007), Heckman and Urzua (2010) and Keane (2010).

This paper contributes to the literature by addressing both of these concerns. It provides a theoretical model illustrating both mobility effects and incentive effects corresponding to each of the two possible directions of the relationship between home ownership and unemployment. Based on this model, I estimate home ownership and job search decisions jointly and adopt state-level mortgage interest deductions as the instrumental variable. Winkler (2010) is the only that paper which adopts the structural approach to estimate the effects of home ownership on labor market outcomes. However, he models the migration decision instead of the job search decision, which does not allow for the investigation of unemployment durations. The estimation strategy in this paper follows the structural estimation of the partial equilibrium job search model (See Burdett and Mortensen (1980), Flinn and Heckman (1982), Eckstein and Van den Berg (2007) and Keane and Todd (2010)).

2.3 Data and Empirical Overview

The primary data in this paper are from the Survey of Income and Program Participation (SIPP) 1996 panel, which is combined with the state-level mortgage subsidy rates provided by NBER. This section starts by describing the two datasets and then presents the empirical evidence on the relationship between home ownership and labor market outcomes. Section 3.3 provides a statistical description and shows that owners tend to be the group with the lower unemployment rate, higher monthly wage and shorter unemployment duration. Section 3.4 adopts a two-stage least-squares approach using the mortgage subsidy rate as the instrumental variable to provide evidence of the causal effect of home ownership on unemployment. Section 3.5 motivates the structural model by summarizing its benefits.

2.3.1 Data

SIPP 1996 Panel

The Survey of Income and Program Participation (SIPP) 1996 panel covers 1996-1999 and collects data every four months (12 periods in total). Each period provides comprehensive information on demographic characteristics and labor force activities, including earnings, number of weeks worked or unemployed, and therefore, a complete history of employment transitions (i.e., transitions from unemployment to employment, or from employment to unemployment) over the interview period. The relatively large sample size[7] and the short recall period[8] make SIPP appealing for studying the dynamics of employment status because it is able to measure relatively short-term unemployment spells.

Meanwhile, SIPP identifies 97 of the largest metropolitan areas with metropolitan codes,[9] which makes it possible to define and track down "movers" by comparing origin and destination metropolitan areas between two consecutive periods. In particular, transition from unemployment to employment with local jobs is defined as unemployed workers ending up with jobs within the same metropolitan area, while a transition from unemployment to employment with NON-local jobs is defined as unemployed workers ending up with jobs in different metropolitan areas. Finally, the survey period, 1996-1999, is associated with a relatively stable housing market, which makes the exclusion of housing price fluctuation and mortgage default a reasonable assumption in the theoretical model. All monetary variables are converted to 1996 dollars.

The sample used in the empirical work is selected from the original sample that contains only white male[10]household heads between 25 and 55 years old. Since

---

[7]The original SIPP 1996 panel has 3,658,293 person-month observations.

[8]NLSY collects data once per year and PSID currently collects data every other year.

[9]Areas are classifies as "non-metropolitan" if the household is located outside a metropolitan area or the metropolitan area is small.

[10]I focus exclusively on males for two reasons. First, it avoids modeling complications due to benefit eligibility for females. Second, job search models are usually used to model the labor force activities of males with high-frequency data, while labor force participation models are used to model females' labor market behaviors.

home ownership and employment transitions are the main variables in this paper, I also drop observations whose residential status are missing; those enrolled in school or the Armed Forces; and those who are self-employed, disabled, retired or not participating in the labor market. The final sample is an unbalanced panel[11] containing 38,112 observations of 4,648 unique individuals.

Mortgage Tax Subsidies

Federal and state income tax policies affect the cost of home ownership. This paper focuses on the mortgage interest deduction, which is the most important favorable tax treatment for home owners. To measure this deduction, I calculate the tax saving from an additional dollar of mortgage interest, the mortgage interest subsidy, based on NEBR publicly available data on tax rates.[12] There are large differences in this tax subsidy across different states. Some states, such as Florida, Nevada and Texas, collect no personal income tax at all, while others, such as California, Delaware, Maine, Massachusetts and North Carolina, rely heavily on personal income taxes to raise revenue, but permit the deduction of mortgage interest. Among these states, the mortgage subsidy rate varies considerably, reaching a maximum of around 9% per dollar of mortgage interest in the District of Columbia. A full list of mortgage subsidy rates in each state between 1996-1999 is provided in Appendix Table A1. The correlation between the state-level mortgage interest deduction and other state-level variables, including annual income per capita, unemployment rate and housing prices, is not significant. The comparison of states with and without the mortgage interest deduction also shows no significant difference in income, household

[11]This sample reduction occurred due to the normal survey attrition, such as refusing to continue to participate, inability to locate persons, deaths, etc.

[12]See details in Feenberg and Coutts (1993) and at http://www.nber.org/taxsim.

characteristics or labor market outcomes. Besides, the deduction variation across states is usually caused by federal and state tax laws.[13] These facts suggest that the mortgage interest deduction is exogenous to individuals' labor market decisions.

One concern about using the mortgage interest deduction as an exclusion restriction in the estimation is whether it is an effective tax policy for promoting home ownership, which is still an open question in the literature. For example, Rosen (1979) and Glaeser and Shapiro (2003) find opposite results: while the former finds the deduction do affect home ownership, the latter suggests this effect is limited. The main argument in the latter is that the home mortgage deduction disproportionately benefits the wealthy because they claim most of the deductions.[14]. The problem is less severe in this paper because it uses the sample of white male household heads, who are more likely than other demographic groups to be affected by this tax policy. Actually, as seen in Section 3.4, reduced-form regressions show that the mortgage interest subsidy is positively associated with individuals' propensity to become homeowners.

2.3.2 Empirical Overview

Owners and Renters as Two Groups

Table(2.1) lists descriptive summary statistics for homeowners and renters as two groups. To emphasize the effect of owning-related costs, I define owners as

[13]Different states implement different formulas for taxable income; some use federal adjusted gross income a starting point for developing their tax base, while others use federal taxable income. And the taxable income in other states is computed independently of the federal formula.

[14]The home mortgage deduction benefits only those households that itemize on income tax returns. That is, the beneficiary's tax liability would be lower when itemizing than when claiming the standard deduction. That is, those who itemize are usually wealthier than those claim the standard deduction. The TAXSIM-based imputed itemization rate is 63.1%.

those with a mortgage, excluding outright owners. [15] As Table(2.1) shows, homeowners, as compared to renters, are older, more likely to be married with child(ren) and have a college education. Labor market outcomes are measured in three dimensions: monthly wage, unemployment rate and length of unemployment spells. Unemployment spells are categorized into two types: those ending with local jobs and those ending with non-local jobs. On average, owners are less likely to be unemployed, have a shorter unemployment duration when unemployed and higher wages when employed. The sample has 951 unemployment spells in total. The average length of unemployment spells for owners is 14 weeks, about one and a half weeks shorter than that for renters.

I pay close attention to unemployment duration. Figure(2.1) compares the Kaplan-Meier survival functions of owners and renters in both local and non-local labor markets. It shows that owners have a higher probability of leaving unemployment to the local market and a lower probability of leaving unemployment to the non-local market, with a higher overall hazard rate (transition probability from unemployment to employment) than that for the renters group.

Lifetime Patterns

Figure(2.2) presents the unemployment rate and monthly wage of owners and renters by age. Renters' unemployment rate has greater volatility than that of owners. With a few exceptions at young ages (< 25) and older ages (around age 48), owners, on average, always have a lower unemployment rate than renters, even

[15] Homeowners with or without mortgages behave differently in the labor market because outright homeowners are more likely to be seniors who have already retired from the labor market(Flatau, Forbes, and Hendershott (2003)) Also, the number of outright owners in the data is much smaller than that of owners who are still in debt.

though the unemployment rates for both groups are quite low (mostly below 10%). Both renters and owners have hump-shaped monthly wages. Owners' wages reach the highest point around age 50, while renters' wages reach the highest point around age 40 and remain flat after that. Through their lifetime, owners, on average, have higher monthly wages than renters.

2.3.3 Reduced-form Estimation

I conduct two-stage least-squares regressions[16] of unemployment on the homeowner indicator using the state-level mortgage interest subsidy as the instrumental variable. Table(2.2) presents the first-stage results. The first column presents results for pooled OLS, while the second and third columns present results for fixed-effect OLS and random-effect OLS, controlling unobserved heterogeneity across individuals. The mortgage interest subsidy is significantly correlated with the ownership indicator in all three models, suggesting that people are more likely to own in the states in which tax policies are more favorable to owners.

Table(2.3) summarizes the regression results of unemployment on the homeowner indicator with or without the instrumental variable. The causal effect parameter is consistently (significantly) negative across different models, even though the magnitude is smaller after adopting the instrumental variable. The negative coefficients show that owners are less likely to be unemployed. The results are consistent with those in Coulson and Fisher (2009), which adopts a probit model for the second stage. The coefficients for the other demographic variables have signs consistent with prior expectations. [17] For example, the ownership equation shows that people

[16] The detailed regression equations are provided in Appendix A2.3.

[17] The estimation of the same model where housing prices are added as regressors show the same patterns of the relationship between home ownership and unemployment.

who are married with children are more likely to be owners. And the unemployment equation shows that people with more education are less likely to be unemployed.

2.3.4 The Benefits of A Structural Model

The empirical evidence so far presents a general picture of how owners and renters are different in terms of labor market outcomes and hints a causal link between home ownership and unemployment. Though the reduced-form work is an efficient method for the purpose of understanding the "effects of causes",[18] a structural framework, which is explicit about how models of counterfactuals are generated, is able to study the "causes of effects" and answer a wide range of questions.

First, the model demonstrates the interaction between home ownership and job search decisions and clarifies the mechanism behind the causal effect. Even though the mobility story implies that owners, as suggested in the data, are less likely to move for a non-local job, no clear explanations are provided for why owners and renters behave differently in local job markets. The model in this paper incorporates both the "mobility effect" and the "incentive effect" and provides explanations for the observed empirical evidence. Furthermore, estimating structural parameters make it possible to test the role of moving costs and the labor market environment behind this causality by simulating ex ante renters and owners with counterfactual parameters. This is hard to incorporate into a reduced-form work because moving costs and certain job offer distributions cannot be directly observed in the data. Second, by explicit modeling the self-selection process of becoming homeowners and the unobserved correlation between home ownership preferences and skill endowment, the model provides alternative identification assumptions to exclude the selection bias from the causal effect. Third, a structural model makes it possible to evaluate the la-

[18]In the terminology of Holland (1989).

bor market consequences of housing policies in new environments and to forecast the effects of new policies. This paper considers the policy of eliminating the mortgage interest deduction and compares its effects in a thriving or a struggling economy. In the remainder of the paper, I develop a model that is rich enough to achieve all these goals and put enough structure on the data to enable estimation of the key parameters governing home ownership and employment transitions.

2.4 Model

In this section, I present an agent's lifetime home ownership and employment transitions in a partial equilibrium model. At each period, the agent decides whether to own or rent, conditional on current home ownership and employment status. When the agent is unemployed, he receives offers from local or non-local labor markets and then decides whether to accept the offer or stay unemployed.

The theoretical model serves two purposes. First, it provides a theoretical base for the analysis of how owning a house affects people's job search behaviors, both in local and non-local labor markets. On the one hand, to avoid moving costs, owners are more likely to accept lower wages in the local market and get out of unemployment in the market sooner than renters (incentive story). On the other hand, to accept a non-local job, owners might require a higher wage to compensate for their moving costs. That is, they are less likely to leave unemployment for a non-local job (mobility story).

Second, the model has implications for the factors that determine people's home ownership decisions. These factors can be divided into three categories. (1) Utility of owning: This preference heterogeneity is usually unobserved, though somehow related to some observed characteristics. For example, in general, a married couple with children is more likely to be owners given that owned houses are more likely to provide a stable living environment; (2) Relative cost of owning: The difference in financial obligations of these two alternatives explains why employed and wealthier people are more likely to become owners. This paper adopts the state-level mortgage interest deduction as the instrumental variable to approximate this cost. It directly affects the home ownership decision, but not the job search decision; (3)Labor market prospects and skill endowment: People with a better and more stable labor market perspective are more likely to meet their financial obligations and purchase, which leads to the endogeneity problem of estimation for the causal effect.

In this section, I first present the binary home ownership decision, specifying its utility and costs, followed by a discussion of the labor market environment and job search decisions. Then, I use Bellman equations to describe how agents make the two sets of decisions simultaneously, as well as the interactions between them. Finally, I discuss the solution method of this dynamic model.

2.4.1 The Decision to Own or Rent

I model the home ownership decision h_t as a binary variable with an additive linear utility u^h. h_t equals 1 when the agent lives in an owner-occupied house and 0 when the agent lives in a rented house.[19] The agent decides whether to rent or to own by weighting the utility and costs of these two alternatives.

Period Utility and Owning Utility

In each period, the agent obtains utility from consumption c_t and a utility premium u^h for being a homeowner. The utility premium u^h captures various

[19]In the housing demand literature–e.g.,Campbell and Cocco (2007) and Yao and Zhang (2005)–housing is usually modeled as a continuous variable (durable good) for the purpose of estimating the substitution elasticities between durable and non-durable goods consumptions.

reasons why people choose owning over renting: e.g., stable living environment, receiving secure tenure, and management control over their dwellings.[20] To capture the observed and unobserved heterogeneity, the utility premium u^h is modeled as the following linear equation:

$$u_t^h = \beta_h X_t + \eta_t^h,$$

where X includes a constant, agent's marriage status, whether the agent has a child(ren) and agent's education level,[21] and η^h is a random variables, $\eta^h \sim N(0, \sigma_{\eta^h}^2)$, which captures the unobserved characteristics that affect owning preferences. Thus, I can define a current period utility function, $v(c_t, h_t)$, which is a function of consumption and the own-or-rent choice:

$$v(c_t, h_t) = \begin{cases} u(c_t); \quad \text{if} \quad h_t = 0 \\ \\ u(c_t) + u_t^h; \quad \text{if} \quad h_t = 1, \end{cases}$$

where $u(\cdot) = \frac{c_t^{1-\rho}}{1-\rho}$, and ρ is the coefficient of relative risk aversion.

Owning and Renting Costs

Owner-occupied housing costs usually consist of a complicated payment schedule, including down payment, monthly mortgage and property interest. Though the terms of renting costs are simple, there is no easy way to incorporate those terms structurally into the theoretical model. Instead, I abstract the costs related to home

[20] Some other determinants are also important when individuals make housing purchase decisions. One is the investment benefit of owning a house, which requires modeling housing-market shocks. This paper omits housing price uncertainty to maintain tractability.

[21] To keep a relatively small state space, I consider whether the agent has a child(ren) instead of tracking the number of children. This is a reasonable simplicity since the number of children might not critical to the ownership choice, though it is an important factor in determining the size of the house that the agent buys.

ownership as two terms. The first is the owning cost ξ^{c}[22], which captures the cost of housing services in the owner mode relative to the cost in the rental mode.[23] The second is moving cost ξ^{m}[24] that is required if owners change their ownership status. This is cost capture owners' inflexibility of changing their ownership status. Unlike renters, who can adjust their housing expenditures more flexibly without a large financial loss, owners commit to a long-term contract, the adjustment of which is costly. This cost is especially large when the housing prices go through a downward trend or, even worse, when the house is under water.

The moving cost is triggered in two cases. In the first case, the owner moves for a new job opportunity and has to let go of her current house. Let binary variable $m_t = 1$ denote the event of moving for a job, while $m_t = 0$ denotes otherwise. Then, this case can be expressed as $(h_{t-1} = 1, h_t = 1, m_t = 1)$ or $(h_{t-1} = 1, h_t = 0, m_t = 1)$, depending on whether the agent chooses to own or rent in the new period. Moving for a job m_t is not an choice variable and is determined by the exogenous labor market environment–that is, whether the unemployed agent receives a non-local job offer and accepts it. In the second case, the agent adjusts her home ownership status without labor market motivations. Similarly, this case can be expressed as

[22]To replicate the payment schedule in the reality, the owning cost can be break down into two terms, one is a one-time fixed cost, and the other is the cost that needs to be paid every period. The one-term owing cost is kept here because it generate the similar home ownership patterns with the model with two-term owning cost.

[23]In studies of home ownership, the annual cost of housing services in the owner mode is generally approximated as the user cost of housing. The user cost is the sum of depreciation and maintenance costs, the after-tax opportunity cost of down payment, the after-tax mortgage interest payment and the after-tax property tax payments minus the expected, nominal capital gain on the housing structure.

[24]The moving costs usually include utility loss of relocation that is hard to measure only by the transaction costs of selling a house, such as fees for brokerage, financing, and title changes.

$(h_{t-1} = 1, h_t = 0, m_t = 0)$.

Letting $\phi(h_t, h_{t-1}, m_t; \xi^c, \xi^m)$ denote the housing-related cost, the above three cases can be summarized as:

$$\phi(h_t, h_{t-1}, m_t; \xi^c, \xi^m) = h_t \xi^c + \begin{pmatrix} \mathtt{1}(h_{t-1} = 1, h_t = 0, m_t = 0) + \\ \mathtt{1}(h_{t-1} = 1, h_t = 0, m_t = 1) + \\ \mathtt{1}(h_{t-1} = 1, h_t = 1, m_t = 1) \end{pmatrix} \xi^m,$$

where $\mathtt{1}$ is a indicator function. At a certain period t, the home ownership cost ϕ is determined not only by the previous and current home ownership status (h_{t-1}, h_t), but also by whether the agent moves for a new job (m_t).

Both the owning cost ξ^c and the moving cost ξ^m have important implications. With the owning cost $\xi^c > 0$, employed people are more likely than unemployed people to become owners, which causes the endogeneity problem. To address this problem and introduce the exogenous variations of home ownership decisions, I approximate this cost by the following linear equation:

$$\xi_t^c = \beta_z Z_t + \eta_t^z = \beta_{z0} + \beta_{z1} Z_{mt} + \eta_t^z,$$

where Z_{mt} is the variable of "mortgage tax subsidy," and η_t^z is the error term that captures the unobserved factors related to the owning cost, $\eta^z \sim N(0, \sigma^2_{\eta^z})$. A higher mortgage interest deduction makes houses more affordable and therefore people are likely to become home owners. Therefore, the variation in this tax code among different states provides identification power in the estimation. I assume away the effects of housing price dynamics on housing costs for simplicity. This assumption is justified given that the data period in this paper covers 1996-1999, during which the housing market was stable. [25]

[25] According to the Federal Housing Financial Agency, the annual housing price, estimated using sales price, increased by about 4.5% from 1996 to 1999. It has to be admitted that housing price

Moving costs capture the inflexibility of home owners in the labor market. First, to compensate for the cost of moving, unemployed owners expect a higher-wage job in the non-local market. Second, to avoid moving costs, unemployed owners are more likely to focus on the local job market, where the new job doesn't require a change in home ownership status. As expected, owners would be more willing to accept a less-satisfying local job simply for the reason that it is local. However, when unemployed, owners might want to become renters to enjoy the geographic freedom, but this option is also impeded by the same moving cost. That is, owners cannot gain back freedom in the labor market by freely adjusting their home ownership status.

Within this simple setup, the home ownership utility premium, together with housing-related cost parameters (u^h, ξ^c, ξ^m), largely determine the agent's decision to own or rent. The idea is quite straightforward: A higher utility premium u^h is associated with a higher home ownership rate, while higher costs (ξ^c or ξ^m) are associated with a lower home ownership rate.

2.4.2 Labor Market and Job Search Decisions

I extend the standard job search model into a two-market model in which an unemployed worker receives a wage offer from not only the local market but also the non-local market, with probability λ_l and λ_n,[26] respectively. The renter can move between the two markets freely, while owners have to pay moving cost ξ^m if they

---

fluctuation would affect the expected and realized owning cost; therefore, it would be an interesting extension to incorporate the housing market dynamic into this paper.

[26]In some job search literature, the probability of a job offer is also determined endogenously by modeling job search intensity. Without the observations on how many offers have been received, it's hard to identify both job search intensity and the probability of accepting a certain wage at the same time. Therefore, I model only the latter as a choice, while keeping the former as exogenous parameters.

accept a non-local job, as mentioned in the last section. This paper models only job-related migration and does not consider the case in which the agent changes locations without changing employment status. Accepting a non-local job is associated with a moving status that is denoted by m_t[27]. That is, $m_t = 1$ when the agent accepts a non-local job. Due to moving costs, the agent responds differently to offers coming from the the two markets. Intuitively, owners are more likely to stay in the local market and are more hesitant when it comes to a non-local job offer.

Wage offers are drawn from the cumulative distribution function $F(w)$. To capture the life-cycle feature of wage growth, I allow wage function to be a quadratic form in age. For simplicity, as in Blau (2011), the wage process depends on age instead of on length of work experience because tracking work experience adds one dimension to the state space. Letting k denote age, the wage equation takes the following form:

$$\ln w(k) = \ln w_{net} + \alpha_1 k + \alpha_2 k^2,$$

where w_{net} is the wage without age accumulation and follows a truncated log normal distribution:

$$\ln w_{net} \sim N(\mu^w, \sigma^{w2} | \ln \bar{w}, \ln \underline{w}).$$

To capture the heterogeneity in the agent's ability or skill endowment, I assume that the mean of wage distribution μ^w is determined by the following equation:

$$\mu_t^w = \beta_w X_t + \eta_t^w,$$

where X includes a constant, the agent's marriage status, whether the agent has a

[27] Instead of directly modeling the destinations of migration choices, as in Gemici (2007) and Winkler (2010), with different characteristics in different locations, the migration status in this paper is determined by job search choice together with the exogenous labor market environment without considering different characteristics in different locations, such as different level of expected labor income.

child(ren) and the agent's education level, and η^w captures the unobserved determinants of the agent's wages with the distribution $N(0, \sigma^2_{\eta^w})$. As in the standard job search model, the wage offer assumption implies that ex ante identical workers may receive different wage offers or, analogously, that the same unemployed worker may receive different offers over time.

For each offer, the unemployed worker decides whether to accept or reject it. While unemployed, the agent receives unemployment benefit b.[28]For the employed worker, the layoff probability is δ. Let e_t denote the working status in period t–that is, $e_t = 1$ when the worker is employed and $e_t = 0$ otherwise. Then, the agent's income I_t is given as

$$I_t = (1 - e_t)b + e_t w_t.$$

For simplicity, I don't consider on-the-job search.[29]

2.4.3 Budget Constraints

Budget constraints reflect the interaction between home ownership and job search decisions in the sense that housing-related cost $\phi(h_t, h_{t-1}, m_t; \xi^c, \xi^m)$ and current income $I_t = (1 - e_t)b + e_t w_t$ simultaneously determine the level of current consumption. To be more specific, the budget constraint is given by:

$$\begin{aligned} c_t &= I_t - \phi(h_t, h_{t-1}, m_t; \xi^c, \xi^m) \\ &= (1 - e_t)b + e_t w_t - \phi(h_t, h_{t-1}, m_t; \xi^c, \xi^m)^{30}. \end{aligned}$$

---

[28]Unemployment benefits net of the cost of search.

[29]The model could be extended to incorporate on-the-job search while the main implications still hold. An interesting direction of further work might be to investigate how home ownership affects workers' job mobility.

[30]Though liqudity constraint can affect the home ownership as well as job search decisions, to avoid making the state space too large and computationally infeasible, the strucutral model doesn't

That is, employment status e_t determines the level of available financial resources, while previous and current home ownership status, as well as current moving status (h_t, h_{t-1}, m_t), determine how much money is left for consumption. I do not formally model the agent's optimal consumption/savings decision. Rather, consistent with much of the previous research in the dynamic, discrete choice literature, I assume that the agent consumes all the available financial resources in each period. [31]

2.4.4 The Household's Choice Set

The agent lives for T ($T > 0$) periods and maximizes expected lifetime utility by making home ownership and job search decisions.[32] The agent's problem is summarized as follows:

$$\max_{\{e_t,h_t\}_{t=0}^T} \sum_{t=0}^{T} \beta^t \mathrm{E}_t[u(c_t) + h_t u_t^h]$$

$$= \max_{\{e_t,h_t\}_{t=0}^T} \sum_{t=0}^{T} \beta^t \ \mathrm{E}_t[u(I_t - \phi(h_t, h_{t-1}, m_t; \xi_t^c, \xi^m)) + h_t u_t^h],$$

where E_t stands for expectation at time t; β is the standard time discount factor; I_t is current income equals to $(1 - e_t)b + e_t w_t$; ξ_t^c is the owning cost depending on housing policy Z_t and equals $\beta_z Z_t + \eta_t^z$; ξ_t^m is the moving cost parameter; and u_t^h is the owning utility premium depending on observed demographic characteristics X_t and is equal to $u_t^h = \beta_h X_t + \eta_t^h$. Each period, the agent observes the realization of

---

incoporate this constraint.

[31]In recent work, researchers have begun to introduce savings into job search models (e.g., Rendon (2006)). Doing so adds a continuous state variable and continuous choice variable into the dynamic model, which significantly complicates the computation. While it is possible from a conceptual standpoint, in practice, this change would make my model intractable at its current level of approximation quality.

[32]This paper is particularly interested in short-term unemployment transitions, which are better captured in job search models instead of in models of labor participation or labor supply.

the three shocks $\{\eta_t^h, \eta_t^w, \eta_t^z\}$, and decides whether to rent or own by weighing the benefit u^h and the cost $\phi(h_t, h_{t-1}, m_t)$, while also taking the employment status e_t into consideration. At the same time, when unemployed, the agent decides whether to accept or reject an offer, local or non-local.

Thus, the agent faces two types of discrete choices: home ownership and job search choices, [33] which give rise to four mutually exclusive feasible choices, $\{d_{it}^{eh}, d_{it}^{er}, d_{it}^{uh}, d_{it}^{ur}\}$, where the first superscript refers to the employment choice and the second to the home ownership choice. Then, the feasible choice set D_t can have the following two forms, depending on the the exogenous labor market environment:

- When an unemployed agent receives a local offer or a non-local offer (4 choices):

$$D_{it} = \{d_{it}^{eh}, d_{it}^{er}, d_{it}^{uh}, d_{it}^{ur}\}.$$

- When an unemployed agent receives no offer or the agent is employed (2 choices):

$$D_{it} = \{d_{it}^{uh}, d_{it}^{ur}\}.$$

In the last two cases, the agent does not have a chance to accept or reject a job offer and only needs to decide whether to own or rent.

2.4.5 State space

The state space S_t consists of all factors known to the agent that affect current utility or the probability distribution of the future rewards. These variables

[33]Though both choices are modeled as binary variables, they are different in terms of the source of randomness. The randomness of the home ownership choice comes from the stochastic term η_h within utility premium u^h as in discrete choice structural models, while the randomness of job search choice comes directly from the stochastic term η_w inside of wage distribution, as in the job search literatures. Both error terms affect the agent's optimal choices.

can be classified into three categories according to their law of motion. The first group includes variables determined by decisions made by the agent in period t: labor market status (e_{t-1}), home ownership status (h_{t-1}) and wage (w_t) if employed. The last also depends on the labor market environment (the available wage offers in the market) because the agent can only decide to accept or reject a wage offer. Specially, wage transitions takes two forms, depending on the current period's employment status. For simplicity, I omit age accumulation here. So, the following w_t and w_{t+1} are the wages without age accumulation.

If the agent is unemployed at period t, then $w_t = 0$ and

$$w_{t+1} = \begin{cases} 0, \quad \text{if the agent receives no job offer or rejects the received job offer;} \\ w_{t+1} > 0, \quad \text{if the agent receives a job offer and accepts it,} \end{cases}$$

where w_{t+1} is a random draw from wage distribution $F(w)$ at period $t+1$.

If the agent is employed at period t, then $w_t > 0$ and

$$w_{t+1} = \begin{cases} w_{t+1} = w_t, \quad \text{if the agent is not laid off;} \\ 0, \quad \text{if the agent is laid off.} \end{cases}$$

The second group of state variables includes those that evolve deterministically, such as three demographic characteristics X_t and two housing policy variables Z_t, where X_t includes whether the agent graduated from college X^e, whether the agent has married X^m, and whether the agent has a child(ren) X^c. Most of those variables are time-variant except for the education variable. Marriage and fertility is exogenous, so in each period, the agent is married or has a child(ren) with a certain probabilities:

$$X^m_{t+1} = \begin{cases} 1 \quad \text{with probability } p_1, \text{ if } \quad X^m_t = 1; \\ 1 \quad \text{with probability } p_0, \text{ if } \quad X^m_t = 0, \end{cases}$$

$$X^c_{t+1} = \begin{cases} 1 \quad \text{with probability 1, if} \quad X^c_t = 1; \\ \\ 1 \quad \text{with probability } p_c \text{, if} \quad X^c_t = 0, \end{cases}$$

The third group of state variables includes three stochastic elements: the home ownership utility stochastic term η^h; the labor market ability stochastic term η^w (affecting the mean of wage distribution); and the owning cost stochastic term η^z. I assume that the three error terms follow a multi-normal distribution $G(\eta^h, \eta^w, \eta^z)$; that is,

$$\begin{pmatrix} \eta^h \\ \\ \eta^w \\ \\ \eta^z \end{pmatrix} \sim N[\begin{pmatrix} 0 \\ \\ 0 \\ \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\eta^h} & \sigma_{hw} & \sigma_{hz} \\ \\ \sigma_{hw} & \sigma^2_{\eta^w} & \sigma_{wz} \\ \\ \sigma_{hz} & \sigma_{wz} & \sigma^2_{\eta^z} \end{pmatrix}].$$

Then, η's have a general contemporaneous correlation structure, but are mutually serially independent. The endogenous problem in the paper comes from the fact that $cov(\eta^h, \eta^w) = \sigma_{hw}$ might not equal to zero.

To sum up, the state space can be expressed as

$$S_t = \{e_{t-1}, h_{t-1}, w_t, X_t, Z_t, \bar{S}_t\},$$

where $\bar{S}_t$ stands for the stochastic part of the state space $\bar{S}_t = \{\eta^h, \eta^w, \eta^z\}$.

2.4.6 Bellman equations

The value function $V_t(S_t)$ can be written as the maximum over the alternative-specific value functions:

$$V_t(S_t) = \max\{V^{uh}_t(S_t), V^{ur}_t(S_t), V^{eh}_t(S_t), V^{er}_t(S_t)\},$$

each of which obeys the Bellman equation. We can think of the optimization process as two steps. In the first step, the agent makes the employment decision, taking the ownership decision as optimal; and in the second step, the agent chooses between two home ownership statuses conditional on the optimal employment decision in the first step. The values of the two employment statuses $V_t^u(S_t)$ and $V_t^e(S_t)$ are the maximum taken over two home ownership choices:

$$V_t^u(S_t) = \max_{\{h_t\}}\{V_t^{uh}(S_t), V_t^{ur}(S_t)\}$$

$$V_t^e(S_t) = \max_{\{h_t\}}\{V_t^{eh}(S_t), V_t^{er}(S_t)\}$$

Current utility depends on the current state and decision, especially, on whether the moving cost ξ^m is triggered. For simplicity of expression, I denote an indirect current utility if the agent moves for a non-local job

$$v_{m1}(S_t, h_t) = u(I_t - \phi(h_t, h_{t-1}, m_t = 1; \xi_t^c, \xi^m)) + h_t u_t^h$$

and an indirect current utility if the agent does not move for a non-local job:

$$v_{m0}(S_t, h_t) = u(I_t - \phi(h_t, h_{t-1}, m_t = 0; \xi_t^c, \xi^m)) + h_t u_t^h$$

The only difference between v_{m1} and v_{m0} is whether the moving costs have been triggered.

Given the above notation and following the job search literature, I present the following Bellman equations, which completely describe the agent's dynamic employment transitions of the agent:

$$\begin{aligned}
V_t^{u\hat{h}}(S_t) \quad &= \quad \lambda_l \Bigg\{ v_{m0}(S_t, \hat{h}_t) + \beta \int \max \left\{ \begin{array}{l} E_{S_{t+1}^{-}}[V_{t+1}^e(S_{t+1})|S_t, \hat{h}_t], \\ \\ E_{S_{t+1}^{-}}[V_{t+1}^u(S_{t+1})|S_t, \hat{h}_t] \end{array} \right\} dF(w) \\
&+ \quad \lambda_n \int \max \left\{ \begin{array}{l} v_{m1}(S_t, \hat{h}_t) + \beta E_{S_{t+1}^{-}}[V_{t+1}^e(S_{t+1})|S_t, \hat{h}_t], \\ \\ v_{m0}(S_t, \hat{h}_t) + \beta E_{S_{t+1}^{-}}[V_{t+1}^u(S_{t+1}^u)|S_t, \hat{h}_t] \end{array} \right\} dF(w) \\
&+ \quad (1 - \lambda_l - \lambda_n)\{v_{m0}(S_t, \hat{h}_t) + E_{S_{t+1}^{-}}[V_{t+1}^u(S_{t+1})|S_t, \hat{h}_t]\} \Bigg\} \\
V_t^{e\hat{h}}(S_t) \quad &= \quad \Big\{ v_{m0}(S_t, \hat{h}_t) + \delta\beta E_{S_{t+1}^{-}}[V_{t+1}^u(S_{t+1})|S_t, \hat{h}_t] + (1-\delta)\beta E_{S_{t+1}^{-}}[V_{t+1}^e(S_{t+1})|S_t, \hat{h}_t] \Big\}
\end{aligned}$$

for $\hat{h} = \{h, r\}$ and the expectations are taken over the three stochastic terms.

The first Bellman equation describes three possible labor market transitions when the agent is unemployed:

- With probability λ_l, the agent receives a local offer and then decides whether to accept or reject it by comparing the lifetime utility of accepting

$$v_{m0}(S_t, h_t) + \beta E_{S^-_{t+1}}[V^e_{t+1}(S_{t+1})|S_t, h_t]$$

  and rejecting

$$v_{m0}(S_t, h_t) + \beta E_{S^-_{t+1}}[V^u_{t+1}(S_{t+1})|S_t, h_t].$$

  Reservation wage ${w^l_{t+1}}^*(S_{t+1}|S_t, h_t)$ in this case is defined as a wage that makes the agent indifferent between accepting or rejecting the offer. That is,

$${w^l_{t+1}}^*(S_{t+1}|S_t, h_t) \equiv \Big\{ w_{t+1}|E_{S^-_{t+1}}[V^e_{t+1}(S_{t+1})|S_t, h_t] = E_{S^-_{t+1}}[V^u_{t+1}(S_{t+1})|S_t, h_t] \Big\}.$$

  The agent will accept any local offer with a wage higher than ${w^l_{t+1}}^*$ and reject any offer with a wage lower than ${w^l_{t+1}}^*$.

- With probability λ_n, the agent receives a non-local offer. Similarly, the agent decides whether to accept this offer or to reject it by comparing the lifetime utility of accepting

$$v_{m1}(S_t, h_t) + \beta E_{S^-_{t+1}}[V^e_{t+1}(S_{t+1})|S_t, h_t]$$

  and rejecting

$$v_{m0}(S_t, h_t) + \beta E_{S^-_{t+1}}[V^u_{t+1}(S_{t+1})|S_t, h_t].$$

  Notice that the key difference in this case, compared with the case of the local job offer, is that the agent has to move ($m_t = 1$) by accepting this offer. Define reservation wage ${w^n_{t+1}}^*(S_{t+1}|S_t, h_t)$ as in the first case; we have

$$w_{t+1}^{n}{}^{*}(S_{t+1}|S_t,h_t) \equiv \Big\{ w_{t+1}|v_{m1}(S_t,h_t) - v_{m0}(S_t,h_t)$$

$$+\beta E_{S_{t+1}^{-}}[V_{t+1}^{u}(S_{t+1})|S_t,h_t] = \beta E_{S_{t+1}^{-}}[V_{t+1}^{e}(S_{t+1})|S_t,h_t] \Big\}.$$

Recall that the moving cost ξ^m is positive, implying that $v_{m1}(S_t,h_t) - v_{m0}(S_t,h_t)$ is negative, which makes it costly for owners to accept a non-local job. The agent will accept any non-local offer with a wage higher than $w_{t+1}^{n}{}^{*}$ and reject any non-local offer with a wage lower than $w_{t+1}^{n}{}^{*}$.

- With probability $1-\lambda_l-\lambda_n$, the agent receives no offer and stays unemployed.

The second Bellman equation describes two possible labor market transitions when the agent is employed. In this case, the agent has no need to make a job search decision and will stay in the same position as long as he/she is not laid off.

- With probability $(1-\delta)$, the agent has an expected future utility of being employed $E_{S_{t+1}^{-}}[V_{t+1}^{e}(S_{t+1})|S_t,h_t]$.

- With probability δ, the agent has an expected future utility of being unemployed $E_{S_{t+1}^{-}}[V_{t+1}^{u}(S_{t+1})|S_t,h_t]$.

To summarize, for each unemployed agent, there are three possible employment transitions, depending on whether the agent receives a local or non-local job offer or stays unemployed. Combined with four possible home ownership transitions, there are twelve possible transitions. For each employed agent, there are two possible employment transitions depending on whether one is laid off. Combined with four possible home ownership transitions, there are eight possible transitions. In total, we have 20 possible transitions (Appendix A4 lists all 20 possible transitions.) Compared with the standard job search model, two features stand out based on the Bellman equations. Firstly, reservation wage is a function of home ownership status, which indicates that employment transitions are affected by whether the agent is an owner or a renter. Second, due to moving costs, owners have different transition probabilities from unemployment into local or non-local jobs.

2.4.7 Numerical solution

The model is formulated as a dynamic program and solved numerically by backward recursion on the value function. For each period, I compute numerical approximations of the value functions, which are then used to approximate those of the previous periods. To be more specific, starting with the value functions for the terminal period, for all values of state values, I assume that

$$V_T^u(S_T) = V_T^e(S_T) = 0.$$

In principle, one could model horizon T[34] as the end of the life-cycle with a retirement spell. However, unreported estimation results suggest that changing the terminal value assumption does not substantially alter parameter estimates. The zero terminal value assumption is, thus, maintained for convenience.

At each period $t + 1$, the agent receives draws from the joint distribution from $G(\eta_h, \eta_w, \eta_z)$ and faces an exogenous labor market environment, which includes the offer arrival rate, wage offers from distribution $F_{\eta^w}(w)$ and the layoff rate. The agent chooses the alternative with the largest realized reward by comparing t period value functions for each alternative within the feasible choice set. Current value functions $V_t^u(S_t)$ and $V_t^u(S_t)$ can be calculated given the expected value function of next period $E_{S_{t+1}^-}[V_{t+1}^u(S_{t+1})]$ and $E_{S_{t+1}^-}[V_{t+1}^e(S_{t+1})]$, as illustrated in the Bellman equations.

---

[34]This paper considers individuals between age 20 and age 55. The time unit in the model is four months, which is consistent with that in the data. So the total periods is $T = 35 * 3 = 105$.

Two features complicate the calculation of expected value function. First, with the continuous variable in the state space, it is infeasible to calculate the expected values $E_{S^-_{t+1}}[V^u_{t+1}(S_{t+1})]$ and $E_{S^-_{t+1}}[V^e_{t+1}(S_{t+1})]$ at each point in the state space S_{t+1}. Second, the expected values are a multi-dimensional integral over the stochastic elements S^-_{t+1} whose realization are not know at time t. The use of a multivariate normal distribution, which allows the correlations, implies that closed-form solutions do not exist.

To overcome the computation problem, I follow the simulation and interpolation method in Keane and Wolpin (1994). On the one hand, the value functions are calculated only at a fraction of the state space and interpolated at the remaining points. On the other hand, joint integrations are approximated by a Monte Carlo simulation. The approximated $E_{S^-_{t+1}}[V^u_{t+1}(S_{t+1})]$ and $E_{S^-_{t+1}}[V^e_{t+1}(S_{t+1})]$ can be used to recover the value function $V^u_t(S_t)$ and $V^e_t(S_t)$. I repeat this procedure until value functions have been approximated for all $t = 0, \cdots, T$. The detailed solution algorithm is provided in the Appendix A 2.4.

2.5 Estimation

2.5.1 Permanent Heterogeneity and Initial Condition Problem

The model represents the decision process of one agent. Differences in the behavior of agents with the same initial state variables arise solely due to serially independent shocks to preferences, wages and owning costs. However, behaviors tend to be more persistent than can be captured by observable state variables. Thus, this paper allows for permanent heterogeneity with respect to the parameters. Specifically, I assume that each individual belongs to a type κ, which is unobserved, and the preference and wage mean parameters are allowed to vary with an agent's type. A simple way to incorporate this feature is by adding time-invariant terms that vary with types into the owning preference u^h and skill endowment μ^w; that is,

$$u_t^h = \beta_h X_t + \gamma^{h\kappa} + \eta_t^h$$

$$\mu_t^w = \beta_w X_t + \gamma^{\mu\kappa} + \eta_t^w,$$

where $\gamma^{h\kappa}$ and $\gamma^{\mu\kappa}$ capture the permanent heterogeneity. People with high $\gamma^{h\kappa}$ are more likely to own, while people with high $\gamma^{\mu\kappa}$ are more likely to be employed.[35] For the rest of the paper, I set $\kappa = 2$ and denote $\Gamma = \{\gamma^{h\kappa}, \gamma^{\mu\kappa}\}$ as type parameters.

Since the model permits unobserved types, the likelihood function is no longer separable[36] and initial conditions cannot be ignored, especially when a large proportion of decisions in the data are observed starting from the middle of the life cycle. Thus, initial conditions are not exogenous because they might relate to permanent heterogeneity in preference and wage distribution parameters. Owning preferences could affect whether or not the agent is a homeowner when first observed, and the wage distribution could affect whether or not the agent is employed when first observed. To account for the initial conditions problem, following the method provided in Wooldridge (2005), I assume that the probabilities of the unobserved heterogeneity types[37] can be represented by parametric functions of the initial outcome variables (employment status and home ownership status). The sample likelihood is

[35]This specification is popular in the context of dynamic discrete decision models. See, e.g., Eckstein and Wolpin (1999) and Keane and Wolpin (1997)

[36]If all components of the likelihood function were additively separable, we could still recover behavior parameters without bias because each piece of the likelihood function could be separately maximized.

[37]As proposed inWooldridge (2005), the initial condition problem is solved by making a flexible distribution assumption of unobserved heterogeneity, instead of specifying initial conditions as functions of unobserved heterogeneity. The model identification is not damaged by assuming the distribution of unobserved permanent heterogeneity as function of observed characteristics.

then calculated by summing up the type-specific likelihoods, as illustrated in next section. When shocks to preferences, wage, and cost equations are serially independent, the initial outcomes are exogenous given types. I assume that an individual can belong to one of two unobserved types. The probabilities that an agent belongs to type 1 or type 2 are given by the following logit form:

$$\begin{aligned}\pi^1 &= Prob(Type = 1) = \frac{X_i^0 \beta_\pi}{1 + X_i^0 \beta_\pi} \\ \pi^2 &= 1 - Prob(Type = 1),\end{aligned}$$

where X_i^0 is a vector of variables that includes a constant, a dummy indicating whether the agent is employed at the beginning of the sample, and a dummy for whether the agent is a homeowner at the beginning of the sample. Parameters in the initial condition equation are estimated together with other behavior parameters in the dynamic model.

For certain parameters, structural estimation inside the model is less crucial than others. I fix such parameters exogenously. For example, the agent is assumed to be risk-averse with coefficient ρ fixed at 2 and the rate of discount β fixed at 0.98. The remaining structural parameters

$$\Theta = \{\Theta^L, \beta_h, \beta_z, \xi^m, \Theta^S\}$$

are estimated using the simulated maximum-likelihood method. Θ includes a vector of parameters governing the labor market transitions $\Theta^L = \{\lambda_l, \lambda_n, \delta, \sigma^w, \alpha_1, \alpha_2, \beta_w, b\}$; owning preference and cost parameters β_h and β_z; moving cost ξ^m; and stochastic parameters $\Theta^S = \{\sigma_{\eta_h}, \sigma_{\eta_w}, \sigma_{\eta_z}, \sigma_{hw}, \sigma_{zh}, \sigma_{zw}, \Gamma, \beta_\pi\}$. Together, I estimate 31 structural parameters.

2.5.2 Simulated Maximum Likelihood

The log-likelihood function is the sum of the individual's log likelihood, which is the density for the sequence of observable conditions on the observable S_i, and the parameters Θ, and unobservable heterogeneity is integrated out by summing over different types.

$$lnL(\Theta) = \sum_{i=1}^{N} \sum_{\kappa} \pi^{\kappa} lnL_i(O_{i0}, O_{i1} \ldots O_{it} \ldots O_{iT} | X_i, Z_i, \Theta^{\kappa}),$$

where π^{κ}, the probability of agent i, is observed as type κ and Θ^{κ} denotes type k parameters[38]. The observed variables are $O_{it} = \{h_{it}, E_{it}, w_{it}, m_{it}\}$–that is, home ownership status ($h_{it} = 1$, own or $h_{it} = 0$, rent), employment status ($E_{it} = 1$, employed or $E_{it} = 0$, unemployed), whether accepting a non-local job ($m_{it} = 1$, accept or $m_{it} = 0$, reject) and wage w_{it}. The first two outcomes $\{h_{it}, E_{it}\}$ are directly determined by the home ownership and job search decisions, while the latter two $\{w_{it}, m_{it}\}$ are determined by job search decision together with the labor market environment. X_i and Z_i stand for demographic characteristics and housing policy variables. They are two sets of exogenous variables in the sense that they are independent from all past, current and future values of the stochastic elements $\bar{S_i}$. Individual i's likelihood contribution can be decomposed into product of transition probabilities $P(O_{it+1}|O_{it}, X_{it}, Z_{it}, \Theta)$, which is calculated based on the optimal solution of the dynamic model and integrated out with respect to unobservable $\bar{S_i}$. That is,

$$L_i(\Theta) = \sum_{t=0}^{T} P(O_{it+1}|O_{it}, X_{it}, Z_{it}, \Theta)$$

$$= \sum_{t=0}^{T} \int p(O_{it+1}|O_{it}, X_{it}, Z_{it}, \bar{S_{it}}, \Theta) dG(\bar{S_{it}}),$$

where $\bar{S_t} = \{\eta_t^h, \eta_t^w, \eta_t^z\}$.

[38]Only two parameters vary with types.

As mentioned in the model section, there are five employment transitions and four home ownership transitions within the model (20 possible transitions in total). The calculation of the above transition probability is complicated for two reasons. First, the probability inside the integration has no analytical form that requires simulation. Second, this probability has to be integrated out over the stochastic part of the state space that has a joint normal distribution. Given the value functions of unemployed and employed $V_t^u(S_t^u)$ and $V_t^e(S_t^e)$, as solved in the dynamic discrete model and the specification of joint distribution $G(\bar{S}_t)$, I use the GHK simulator (See Geweke [1991], Hajivassiliou [1990] and Keane [1994]) to approximate the probability expressions in the likelihood function. The BHHH algorithm is used for maximizing the simulated likelihood function. More details on calculation and the maximization of the likelihood function are presented in Appendix A 2.5.

2.5.3 Identification

The identification of the model relies on functional assumptions as well as on exclusion restriction. The mortgage interest deduction, as the exclusion restriction, is an effective tax policy for promoting home ownership, even though it might disproportionately benefits the wealthy because they claim most of the deductions [39]. At the same time, it is not correlated with local level employment status. Those characteristics of this state-level policy make it possible to identify the effect of home ownership on unemployment. What's more, I address the question of identification in two ways.[40] First, I examine a numerical estimate of Hessian at the estimated

[39]See Glaeser and Shapiro (2003)

[40]For more information on the identification of dynamic discrete choice models, see the discussions in Rust (1994),Magnac and Thesmar (2002),Kasahara and Shimotsu (2009),Abbring (2010),and French and Taber (2011).

parameter value and make sure it is nonsingular. Second, I show that each of the parameters has influence on the observed transition and outcome probabilities, which gives an informal argument for identification.

The challenge of identifying labor market parameters Θ^L comes from the fact that rejected job offers are not observed. This problem is solved by a distribution assumption on wage offers.[41] Given log-normality distribution, the wage offer parameters are identified from the observed wages among the population, conditional on their being employed. Wage growth parameters α_1 and α_2 are pinned down by a wage increase (decrease) in two consecutive periods, conditional on working. Wage mean parameters β_w determine the heterogeneity of wage distribution among different demographic groups and are identified by the difference in wages among these groups. The two arrival rates, (λ_l, λ_n), are identified from the probability of working in local or non-local markets, conditional on not working in the previous period. The job lay-off rate δ is identified from the probability of not working, conditional on having worked in the previous period.

Owning preference parameters β_h and cost parameters β_z are identified from the home ownership transition probabilities, conditional on different observed state variables. More specifically, the solution to the optimization of ownership choices in the model provides the rules of those observations. The identification of cost parameters β_z depend on the state-level variation of mortgage interest deduction. Owning preference parameters β_h are identified by the fraction of individuals who switch between two alternative ownership statuses within different demographic groups. Moving cost ξ^m is the key factor that drives owners to behave differently when facing local versus non-local job offers. It is identified mainly from the different transition probabilities of owners from unemployment to local and non-local employment. Meanwhile,

[41] See Flinn and Heckman (1982),Eckstein and Van den Berg (2007) and Abbring (2010).

facing considerable moving costs makes people more cautious when purchasing houses. Therefore, the probability of ownership also contributes to the identification of moving costs.

The stochastic parameters Θ^S determine the number of unobserved factors and heterogeneities for ownership and employment choices. Without those stochastic terms, ownership and employment outcomes and transitions would be driven only by behavior parameters and observed state variables. Therefore, Θ^S are identified by different observations conditional on the same combination of observed state variables. In particular, the covariance of unobserved preference and skill endowment σ_{hw} is identified by the exclusion restriction, which affects only the selection process of becoming a homeowner. In addition, the observed history of the states outcomes not only determines the agent's current state, but is also statistically informative on identifying agent's unobserved persistent characteristics (Γ, β_π).

2.6 Results

2.6.1 Parameter Estimates

The estimate results[42] and corresponding asymptotic standard errors for the three versions of the model are reported in Table(2.4). Asymptotic standard errors

[42] I tried three version of the model for estimation. The first version excludes the effects of demographic characteristics (marriage, child(ren) and education) in the owning utility equation and wage mean equation and does not allow for permanent unobserved heterogeneity. The second version adds demographic characteristics to better proxy unobserved the owning preference and unobserved skill endowment. The third version models permanent unobserved heterogeneity as discrete types to capture the observed persistent behaviors in the data. The third version is the one presented in the paper. The estimation results of the other two versions are quite consistent with the third version. These results are available from the author upon request.

are calculated using the outerproduct gradient estimator.

Labor market parameters $\Theta^L = \{\lambda_l, \lambda_n, \delta, \sigma_w, \alpha_1, \alpha_2, \beta^w, b\}$: First, the probability of receiving a local offer, λ_l, is much higher (around 0.9) than that of receiving a non-local offer, λ_n (around 0.1). The layoff probability, δ, is relatively small, lower than 0.05. Parameters in the wage accumulation equation show a hump shape of wage over age. Rendon (2006) finds similar labor market parameters. In addition, the wage mean equation shows that marriage, child(ren) and education are all positively correlated with the individual's wage level.

Home ownership parameters $\{\beta_h, \beta_z, \xi^m\}$: The utility premium of owning is positively associated with marriage, child(ren) and education, which is consistent with the reduced-form results in Section 3. The owning cost equation shows that a higher mortgage interest subsidy implies a lower owning cost, which would lead to a higher probability of owning. I estimate a considerable moving cost, about \$700, which is quite unaffordable for an unemployed worker whose unemployment benefits are estimated to be less than \$1000. This explains why owners are more eager to get a job in the local market and are reluctant to move. The moving cost estimated in this paper is smaller than those in previous papers, such as Gemici (2007) and Winkler (2010), for two reasons. First, the moving cost here measures the additional cost that the agent has to pay as an owner versus as a renter, not the total moving cost. Second, the geographic unit in this paper is a metropolitan area, which is a smaller unit than the nine regions defined in the two cited papers.

Stochastic parameters $\Theta^S = \{\sigma_{\eta_h}, \sigma_{\eta_w}, \sigma_{\eta_z}, \rho_{hw}, \rho_{hw}, \rho_{zw}, \Gamma, \beta_0\}$: First, the covariance between owning utility and wage mean shocks σ_{hw} is large and significant, which confirms the endogeneity of the ownership decision. Second, the variance of the owning cost equation error η_z is quite large, which is reasonable since I adopt only one variable (mortgage interest subsidy) to approximate the relative cost of owning to renting. In the third estimation model, type parameters show that people with a higher preference for owning also have a higher mean wage, which indicates the existence of permanent heterogeneity. At the same time, considering unobserved heterogeneity not only largely decreases the variance of preference and wage shocks, but also lowers the covariance between these two shocks.

2.6.2 Model selection: adaptive least absolute shrinkage and selection operator(LASSO)

The structural model in this paper has relatively flexible assumptions on the form of owning utility and wage mean equations and, at the same time, allows for the correlation among three error terms $\{\eta^h, \eta^w, \eta^z\}$. This setup means that a relatively large number of parameters need to be estimated.[43] To avoid the potential problem of mis-specification, I conduct a robustness check exercise by adopting the adaptive least absolute shrinkage and selection operator (LASSO) method. It is an innovative model selection method that selects a relatively parsimonious set of most relevant parameters, which is used mainly for linear regression models with a large collection of possible covariates. But the same idea within the model can be generalized and applied to a more structural estimation.

As Tibshirani (1996) proposes, LASSO maximizes likelihood subjects to a penalty function, the sum of the absolute value of the coefficients being less than a constant. This method shrinks some coefficients and sets others to 0. It keeps the most relevant estimation parameters and reduces estimation variance. However, LASSO can be inconsistent for variable selection for certain condition. To avoid this problem, Zou (2006) proposes a modified version of LASSO, called adaptive LASSO,

[43] The number of structural parameters in the model is 31.

where adaptive weights are used in penalty function. That is,

$$\max lnL(\Theta) \qquad s.t. \sum_j |\hat{w_j}\theta_j| \leq t,$$

where $\theta_j \in \Theta$ and t is a constant called the tuning parameter and $\hat{w_j} = 1/|\hat{\theta_j}|^\gamma$ is the adaptive weight and $\hat{\theta_j}$ is the estimated results from simulated maximum likelihood. Adding the adaptive weight guarantee good asymptotical properties of the LASSO solution. A key step of this method is finding good turning parameter and the adaptive weight parameter (t, γ) that controls the regularization incorporated by the inequality constraint. Following the literature, I choose t by minimizing K-fold cross validation error ($K = 5$), an approximation of the mean squared error of the estimation. The algorithm for the adaptive LASSO solution and the K-fold cross validation criterion are presented in Appendix A 2.6.

Tables(2.4) presents the corresponding adaptive LASSO solutions for the three versions of the estimation models. The parameter for age-square (α_2) in the wage accumulation equation goes to reason. A possible reason that this parameter goes to zero is that this paper includes only individuals who are between 20 and 55. Within this age range, people's wages have not yet started to decline dramatically with age yet. Besides, the coefficients for marriage and child(ren) (β_{w1}, β_{w2}) in the wage equation and the coefficient for education (β_{h2}) in the preference equations become zero, which is consistent with the reduced-form results that marriage and child(ren) are usually not significant in the wage equation. Finally, the parameters denoting the permanent ability level (γ^{w1}, γ^{w2}) go to zero, indicating that the data used in the paper might not be able to distinguish between individuals with different unobserved ability.

2.6.3 Model fit

In this section, I show that the model predictions fit the data reasonably well from two perspectives. First, the model simulations, in general, do well in replicating the home ownership rate among the population of different demographic groups. Second, the model simulations capture the differences between homeowners and renters in terms of labor market outcomes and transitions.

Table(2.5) compares the predicted and actual home ownership rate by agents' marriage status, child(ren) and education level. The total ownership rate is around 70 % in both the data and the predictions. As in the data, the predicted home ownership rate is higher among people who are married with child(ren) and have a college education. Groups by marriage status show the largest difference. The ownership rate (81%) among those who are married is almost double that of those who are not married (44%). Figure(2.3) shows that the model is able to fit the home ownership rate over the life cycle. It gradually increases with age because owning utility and monthly wage increase with age. The model under-predicts young people's home ownership rate and over-predicts old people's home ownership rate. The under-prediction can be explained by the lack of a capital market in the model. Without the option of precautionary saving, people are more cautious. They will wait until their labor incomes reach a certain higher level when considering the purchase of a house because the adjustment of owning is costly. So, under-prediction happens among young people whose wage level is relatively low. One reason for over-prediction is that the model assumes that children stay with parents and add to the utility premium of owning throughout the entire life cycle. This assumption reduces the possibility that parents downgrade their housing or switch to renting after their children leave for college or a career, which happens often in reality.

Table(2.6) and Figure(2.4) show that the model is able to fit the wage and employment profiles of homeowners and renters. In Table 10, the model captures the following three data features: (1) homeowners, on average, have higher monthly wages; (2) homeowners have a lower unemployment rate; and (3) homeowners have shorter unemployment spells that end with local jobs, but longer spells that end with non-local jobs. Overall, homeowners leave unemployment faster and have shorter unemployment spells. In addition, Figure(2.4) presents a graphical comparison of actual and predicted wage and employment profiles of homeowners and renters by age. The upper part of Figure(2.4) shows that the model fits the increasing trend of the life cycle wage profile. The lower part of Figure 2.4 shows that the model fits the fact that the unemployment rate is relatively low (mostly below 2%) across the life cycle.[44]

Table(2.7) compares the predicted and actual transition probabilities out of unemployment and employment for owners and renters as two groups. The model fits the transition probabilities reasonably well. As in the data, owners have a higher probability of leaving unemployment with a local job, but a lower probability of leaving unemployment with a non-local job. In both predicted and actual values, the transition probability to local jobs is much higher than that to non-local jobs. For the transition probability from employment to unemployment, owners and renters seem have a relatively small difference(0.12%) in the model, compared with that in the data(2.05%). This is because the job search model in this paper doesn't allow voluntary leave, so the layoff rate is determined mainly by the single layoff parameter δ, which is the same for both owners and renters.

---

[44]This unemployment rate is lower than that in the whole population because I consider only white male workers.

2.7 Counterfactual experiments

After recovering the underlying parameters of the model and assessing their success in replicating the data, I perform four counterfactual experiments. The first experiment quantifies the causal effects of home ownership on the labor market outcomes by simulating two groups of individuals whose home ownership choices are exogenously restricted. The second and third experiments investigate how the causal effects are affected by moving costs and the labor market environment. The fourth experiment studies the labor market consequences of eliminating the tax policy promoting home ownership.

2.7.1 The effect of home ownership on unemployment

Consider two groups of individuals with the same initial conditions and facing the same environment, I simulate their job search behaviors and labor market outcomes. The only difference between the two groups is that individuals in the first group are not allowed to own, while those in the second group are not allowed to rent. In this way, exogenous renters and owners are generated. The effects of home ownership can then be measured by comparing the labor market outcomes of individuals in these two groups without the endogenous bias.

Table(2.8) presents the unemployment rate, monthly wage and unemployment spells for simulated renters and owners. It shows that owners have a lower unemployment rate (-0.17%), lower monthly wages (-33$) and a shorter unemployment spells (-1.32 weeks). Directly comparing owners and renters, owners have lower unemployment rate by about 2% (See Table 2.2), which means that causal effect explains about 8.5%[45] of the difference between those two groups. However, accord-

[45]I calculate this number by dividing 0.17% by 2%.

ing to the random-effect linear regression of unemployment on home ownership with instrumental variable, home ownership decreases the unemployment rate by about 0.74% (See Table 2.3), which is larger than 0.17% and suggests that the linear form assumption biases the causal effect upward.

According to the estimated parameters, owners have to pay a higher cost to enjoy a relatively high utility premium. Thus, they have to work more to maintain the housing service and cover the cost. This is why owners have a lower unemployment rate. Their average monthly wage is lower because they sometimes are forced to accept lower wages to maintain a smoother labor income to pay owning costs, even though the difference in wage is relatively small. The effects of home ownership on unemployment spells that end in local and non-local jobs are opposite, corresponding to the "mobility effects" and "incentive effects." Home ownership shortens unemployment spells that end in the local market by about one and a half weeks, while it prolongs those that end in the non-local market by about two weeks. For renters, the differences in unemployment durations in the two markets are determined solely by the job arrival rates from the two markets because renters are free to move. Owners' job search behaviors, however, are different from renters' in both local and non-local markets. On the one hand, to avoid moving costs, they have a greater incentive to be reemployed in a local job, which leads to short unemployment durations (1.56 weeks). On the other hand, owners would accept and move for a distant job offer only if the wage offered were high enough to cover the moving costs. It usually takes a much longer time for unemployed workers to be matched with such job offers, which leads to longer unemployment durations that end in non-local jobs (1.77 weeks). Overall, owners have shorter unemployment spells because relatively a large proportion of job offers come from the local market, as estimated from the data ($\lambda_l = 0.8, \lambda_n = 0.1$). That is, the incentive effects dominate the mobility effects.

Certain demographic variables are particularly important for the effects of home ownership on unemployment. To see this, I simulate different groups of individuals depending on whether they have college education and whether they are more than 40 years old. Thus, we have four groups: young workers with high education or low education and old workers with high education or low education. Table(2.9) presents the effects of home ownership on unemployment durations for the four groups. It shows that young workers with low education are more likely to be affected by home ownership, remaining unemployed about two weeks longer if live in owned houses. This is a reasonable result since young unemployed individuals with low education are less likely to receive higher wages from a distant job market and, thus, are more likely to stay in the local market. With this expectation, they would be more willing to accept local offers. For similar reasons, home ownership has the smallest effect on old workers with high education.

Unlike the data comparison, where differences in the labor market outcomes of owners and renters can also arise due to endogeneity on observable and unobservable characteristics, the difference in behaviors between simulated owners and renters can be attributed solely to ownership status itself, as the two groups of individuals have the same characteristics and face the same environment. Therefore, the differences between the two simulated groups measure the causal effect of home ownership on unemployment, and, from this, I conclude that home ownership decreases the unemployment rate and unemployment durations.

2.7.2 The Role of Moving Costs

As the model illustrates, the positive moving cost associated with home ownership alters the job search behaviors of homeowners when they face local or non-local job offers. However, this moving cost is hard to measure and cannot be directly observed in the data. So, it is impossible to measure the effects of moving costs in reduced-form regressions. Instead, based on the estimated model, I simulate behaviors under a counterfactual experiment in which the cost of moving is set to half of the estimated value or zero to capture the important role that it plays.

Table(2.10) shows the differences between the group of homeowners and the group of renters in the experiment setup compared with the baseline setup. With smaller moving costs, the effects of home ownership on unemployment are also smaller, though the direction does not change. To be more specific, in this case, home ownership lowers the unemployment rate by 0.08% and decreases unemployment spells by 0.53 weeks. Without moving costs, owners still have a lower unemployment rate(-0.03%) and shorter unemployment spells (-0.04 weeks); however, the differences become quite small, almost negligible. This is because, with zero moving cost, owners become pickier in the local market and are relatively indifferent between local and non-local offers. Both "mobility effects" and "incentive effects" are eliminated when the moving cost goes to zero. Though this paper does not allow for heterogeneity in moving costs, these costs can vary dramatically across populations. For people with strong local ties−−for example, those with a large family who have lived in the local area for several generations−−the cost of leaving the local market is particularly high, and home ownership would have a greater impact when they are unemployed. The downward trend of housing prices is another reason for high moving costs. Especially with underwater houses, owners face a loss of assets that adds up to the moving cost. Thus, the effects of home ownership are larger during housing market downturns. This counterfactual experiment shows that the incentive to get back to work to avoid moving costs is the main driving force behind the causal effects of home ownership on unemployment.

2.7.3 The role of the labor market environment

The net effect of home ownership on unemployment spells depends on the empirical magnitudes of effects on local and non-local labor markets. This section studies the influence of the labor market environment by setting up alternative job offer arrival rates in the two markets. In the baseline model, the total job offer arrival rate is $\lambda_l + \lambda_n = 0.8 + 0.1 = 0.9$. I keep this total the same in the counterfactual experiments, while considering two extreme cases. The first case excludes job offers from the distant labor market ($\lambda_l = 0.9, \lambda_n = 0$), while the second case excludes job offers from the local market ($\lambda_l = 0, \lambda_n = 0.9$).

Table(2.11) shows the differences between the group of homeowners and the group of renters in the two extreme labor market environments. When no offers come from distant labor markets, owners still have a lower unemployment rate (-0.15%) and shorter unemployment spells (-0.53 weeks), but with a smaller magnitude because owners are less worried about the necessity of moving. The mobility mechanism is eliminated in this extreme case, and the difference between owners and renters is caused solely by the incentive effects. In the case in which no offers come from the local labor market, owners are more likely to become unemployed (0.11%) and stay in unemployment for a much longer time (1.59 weeks) than renters. This is because unemployed owners can get back to work only by accepting a job offer with moving costs; that is, the incentive mechanism is shut down and only the mobility mechanism is effective. Therefore, the direction of the effects of home ownership is consistent with those predicted in Oswald's hypothesis. By controlling the job arrival rates from the local or non-local market, this counterfactual experiment shows the opposite direction of incentive and mobility effects. This explains why Oswald's hypothesis gets little empirical support from individual data, even though it provides a reasonable story. This hypothesis makes a strong assumption that a large proportion of unemployed workers have to move to get reemployed, which is not the case in reality.

This counterfactual implies that the role of home ownership in economic recovery varies according to whether the local economy is relatively thriving or struggling. Take two cities, Detroit and Houston, as examples. At the end of the recent recession, both cities have the relatively same level of home ownership rates. The major difference between the two cities is that Detroit's economy depends greatly on the auto industry, which has been hit hard during the recession, while Houston's economy relies largely on the energy and utilities industry, which have been affected less during the recession. The above exercises, therefore, conclude that home ownership may hinder the economic recovery in Detroit, but has a smaller effect on unemployment in Houston.

2.7.4 The labor market consequences of eliminating the mortgage interest subsidy

This section quantifies the labor market consequences of housing policy–in particular, the mortgage interest deduction–by simulating behaviors with no mortgage interest deductions. Table(2.12) presents the outcomes under the baseline and the alternative policy. Without this favorable tax policy, people are less willing to own. The home ownership rate decreases from 69 % to 55 %. Accordingly, the unemployment rate increases by 0.02 %. The change in the duration of unemployment is a combined result of unemployment that ends in local jobs and unemployment that ends in non-local jobs. The local market effects exceed the non-local market effects and lead to unemployment spells that are about three days longer. The policy implication of this experiment is that eliminating mortgage interest is not necessarily an efficient method of decreasing unemployment, though it might be a way to raise more money for the government and ease the budget deficit.

The effects of the same policy vary under different circumstances. To see this, I extend the same policy change into an alternative environment where agents receive half of their job offers from the local market and the other half from the non-local market($\lambda_l = 0.45, \lambda_n = 0.45$). The lower part of Table 14 presents the outcomes under the baseline and the alternative policy in this alternative environment. Similarly, without this favorable tax policy, people are less willing to own. What's more, with an expectation that a higher proportion of offers will come from the distant market, the downside of home ownership prevents more people from becoming homeowners. The home ownership rate decreases from 69% to 52%. Contrary to the above case, eliminating the mortgage interest deductions does decrease unemployment rate by 3% and shorten unemployment durations by about four days. That is, the labor market consequences of this policy depend on the environment in which it is implemented.

2.8 Conclusion

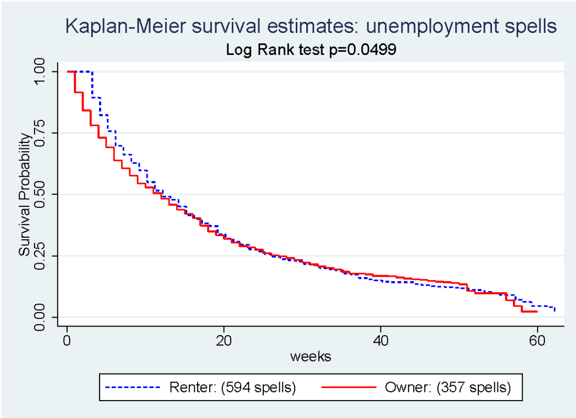
This paper contributes to the literature in two ways. Theoretically, it generalizes the standard job search model to allow for local and non-local job offers and endogenous home ownership decisions. By incorporating these additional features, the model illustrates both the incentive and the mobility mechanisms, which separately explain why homeowners have a higher hazard rate from ending unemployment with local jobs and a lower hazard rate from ending unemployment with non-local jobs. Empirically, this paper estimates the overall causal effects of home ownership on unemployment in a structural approach. To address the endogeneity concern, the model allows for correlation between home ownership and job search decisions and provides an exclusion restriction with exogenous variation of mortgage policies across different states. The estimates show that home ownership decreases the unemployment rate and unemployment duration. That is, incentive effects overtake mobility effects. The counterfactual with zero moving costs and an alternative labor market environment shows that positive moving costs and a relatively large job offer arrival rate from the local market are two main prerequisites for the effectiveness of the incentive mechanism. The counterfactual experiments where mortgage interest subsidy is eliminated show that it affects labor market outcomes differently in different labor market environments.

Table 2.1: Summary Statistics by Ownership status

	Renter[2]	Owner[3]	Total
Age	37.06	41.46	40.31
	(9.42)	(8.43)	(8.91)
Married	51%	85%	76%
	(0.50)	(0.36)	(0.43)
With children	41%	59%	55%
	(0.49)	(0.49)	(0.50)
With College education	52%	63%	60%
	(0.50)	(0.48)	(0.49)
Unemployment	4%	2%	3%
	(0.20)	(0.14)	(0.16)
Wage ($1,000)	2.53	3.92	3.56
	(2.26)	(3.51)	(3.28)
Unemployment duration (weeks)	15.53	13.89	14.71
	(12.40)	(12.80)	(12.62)
End with local jobs (weeks)	14.84	13.35	14.08
	(11.83)	(12.34)	(12.10)
End with non-local jobs (weeks)	17.37	17.56	17.45
	(10.85)	(1.807)	(9.240)
Local jobs	79%	84%	81%
	(0.41)	(0.37)	(0.39)
Non-Local jobs	7%	5%	6%
	(0.26)	(0.21)	(0.24)
Right censored	14%	11%	13%
	(0.35)	(0.33)	(0.34)
No. of Individuals	1,559	3,089	4,648
No. of Observations	9,413	28,699	38,112

[1] Data source: SIPP 1996 Panel (1996-2000). Sample: Ages 20-55, white male household head.
[2] Owner: exclude outright owner;
[3] Renter: exclude those who live in public housing;
[4] Mean of each variable with standard deviation in parentheses.

Figure 2.1: Unemployment Duration with two destinations: Survival functions



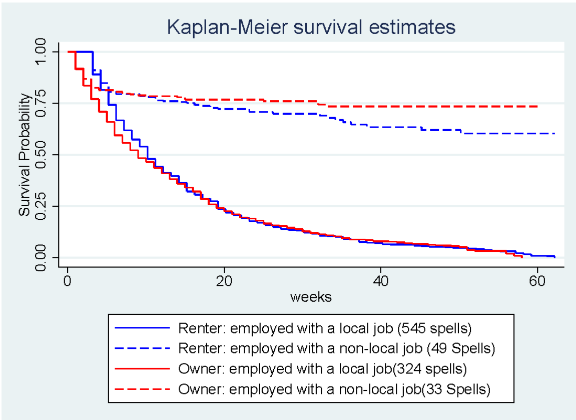
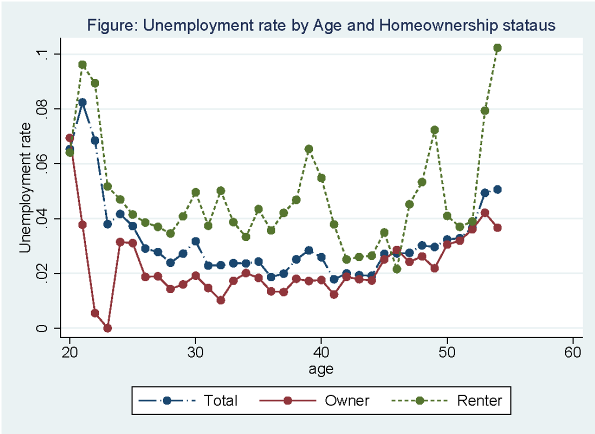


Figure 2.2: Unemployment rate and monthly wage by home ownership status and age



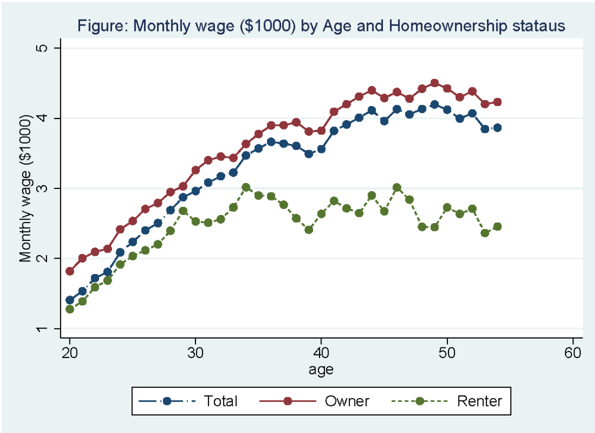


Figure 2.3: Model fit: Home ownership rate by age

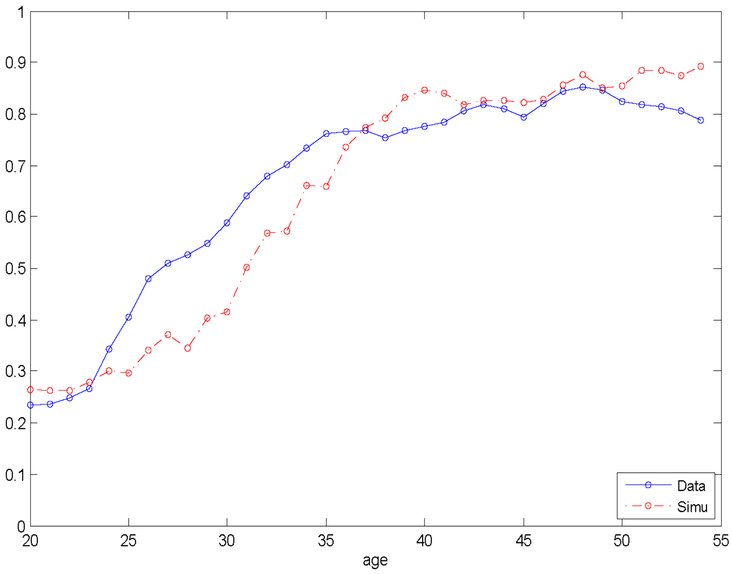


Table 2.2: OLS First stage: Homeownership and Mortgage Subsidy Rate

	OLS	OLS fixed effect	OLS random effect
Mortgage rate subsidy (%)	0.1839***	0.6496***	0.6057***
	(0.0390)	(0.0778)	(0.0732)
Age	0.0111***	0.0115***	0.0124***
	(0.0001)	(0.0010)	(0.0006)
Married (%)	0.2918***	0.1380***	0.1469***
	(0.0028)	(0.0034)	(0.0033)
With children (%)	0.0558***	0.0317***	0.0341***
	(0.0024)	(0.0025)	(0.0025)
With College education(%)	0.0692***		0.0841***
	(0.0021)		(0.0117)
Constant	0.0259*	0.1471***	0.0236
	(0.0108)	(0.0415)	(0.0246)
Observations	38,112	38,112	38,112
R^2	0.1674	0.0365	0.0341

[1] Data source: SIPP 1996 Panel (1996-2000). Sample: Ages 20-55, white male household head;
[2] Year dummies are included in the regressors but not reported in the table;
[3] Standard errors in parentheses;
[4] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.3: Ownership on unemployment: IV Approach

	OLS	IV	OLS FE	FE with IV	OLS RE	RE with IV
Home ownership	-0.2868**	-0.0225***	-0.2406*	-0.0008***	-0.1276	-0.0074***
Status (0,1)	(0.1068)	(0.0010)	(0.1031)	(0.0001)	(0.1028)	(0.0019)
Age	0.0006***	-0.0028*	-0.0015*	0.0011	0.0003	0.0018
	(0.0001)	(0.0012)	(0.0007)	(0.0015)	(0.0002)	(0.0013)
Married (%)	-0.0078***	-0.0980**	0.0000	0.0329*	-0.0044	0.0177
	(0.0011)	(0.0312)	(0.0028)	(0.0145)	(0.0024)	(0.0186)
With children (%)	0.0062***	-0.0110	0.0020	0.0096*	0.0020	0.0073
	(0.0010)	(0.0061)	(0.0021)	(0.0039)	(0.0019)	(0.0048)
With College	-0.0148***	-0.0362***			-0.0194***	-0.0099
education(%)	(0.0008)	(0.0075)			(0.0036)	(0.0092)
Constant	0.0353***	0.0206**	0.0942***	0.1417***	0.0468***	0.0467***
	(0.0036)	(0.0065)	(0.0276)	(0.0400)	(0.0086)	(0.0102)
Observations	38,112	38,112	38,112	38,112	38,112	38,112
R^2	0.0306	0.0188	0.0137	0.0079	0.0249	0.0204

[1] Data source: SIPP 1996 Panel (1996-2000). Sample: Ages 20-55, white male household head;
[2] Year dummies are included in the regressors but not reported in the table;
[3] Standard errors in parentheses;
[4] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2.4: Estimation results (31 parameters, 38,112 observations): with demographic characteristics (marriage,children and education); unobserved heterogeneity (2 types)

		SML		Adaptive Lasso results[1]	
Parameter	**Symbol**	**est.**	**std.dev**	**est.**	**std.dev**
Labor market parameters:					
local job arrival rate	λ_l	0.862	(0.177)	0.734	(0.135)
non-local job arrival rate	λ_n	0.117	(0.009)	0.089	(0.004)
layoff rate	δ	0.033	(0.013)	0.024	(0.008)
variance of log wage distribution	σ^w	0.297	(0.121)	0.240	(0.116)
wage accumulation equation	α_1	0.094	(0.040)	0.079	(0.023)
	α_2	-0.0021	(0.0090)	0	(0)
wage mean equation	β_{w0}	0.521	(0.029)	0.431	(0.015)
	β_{w1}	0.081	(0.049)	0	(0)
	β_{w2}	0.039	(0.044)	0	(0)
	β_{w3}	0.347	(0.181)	0.189	(0.098)
unemployment insurance benefits	b	0.879	(0.162)	0.577	(0.114)
Home ownership parameters:					
owning utility equation	β_{h0}	0.442	(0.153)	0.405	(0.101)
	β_{h1}	0.044	(0.076)	0.012	(0.032)
	β_{h2}	0.776	(0.133)	0.411	(0.104)
	β_{h3}	0.032	(0.052)	0	(0)
owning cost equation	β_{z0}	0.644	(0.176)	0.342	(0.087)
	β_{z1}	-0.421	(0.123)	-0.319	(0.087)
moving cost	ξ^m	0.797	(0.312)	0.506	(0.144)
Stochastic parameters:					
error standard deviation	σ_{η_h}	0.321	(0.143)	0.226	(0.089)
	σ_{η_w}	2.442	(0.512)	2.021	(0.376)
	σ_{η_z}	5.372	(0.121)	4.215	(0.954)
covariance matrix	σ_{hw}	0.217	(0.108)	0.126	(0.046)
	σ_{wz}	0.122	(0.042)	0.102	(0.029)
	σ_{hz}	0.017	(0.012)	0.011	(0.009)
unobserved permanent heterogeneity	γ^{h1}	0.418	(0.010)	0.301	(0.006)
	γ^{h2}	0.137	(0.057)	0.105	(0.036)
	γ^{w1}	0.536	(0.657)	0	(0)
	γ^{w2}	0.132	(0.120)	0	(0)
type probability equation	$\beta_{\pi0}$	0.038	(0.002)	0.029	(0.001)
	$\beta_{\pi1}$	0.842	(0.154)	0.625	(0.121)
	$\beta_{\pi2}$	0.548	(0.132)	0.351	(0.076)

[1] Penalty parameter is chosen by the K-fold cross validation where K=5.

Table 2.5: Model fit: Home ownership rate

	Data	Model
Home ownership rate by marriage status		
Married	81%	79%
Not Married	43%	44%
Home ownership rate with or without child(ren)		
With Child(ren)	79%	75%
Without Child(ren)	64%	65%
Home ownership rate by education		
With College Education	77%	75%
Without College Education	64%	68%
Total	72%	69%

Table 2.6: Model fit: labor market outcomes by home ownership status

	Renter		Owner		Total	
	Data	Model	Data	Model	Data	Model
Monthly Wage ($1,000)	2.53	2.36	3.92	3.64	3.56	3.54
Unemployment rate	4%	4%	2%	2%	3%	2%
Unemployment duration(weeks)	15.53	15.25	13.89	13.56	14.71	14.66
End with local jobs(weeks)	14.84	14.26	13.35	12.38	14.08	13.88
End with non-local jobs(weeks)	17.37	16.98	17.56	18.21	17.45	17.21

Table 2.7: Model fit: labor market transitions by home ownership status

	Renter		Owner		Total	
	Data	Model	Data	Model	Data	Model
Transition from unemployment						
to local jobs	43.84%	44.26%	42.79%	43.61%	43.36%	43.40%
to non-local jobs	3.44%	3.96%	2.29%	3.17%	2.90%	2.94%
to unemployment	52.72%	51.78%	54.92%	53.22%	53.74%	53.66%
Transition from employment						
to unemployment	4.43%	3.21%	2.48%	3.33%	3.22%	3.25%
to employment	95.57%	96.79%	97.52%	96.67%	96.78%	96.75%

Table 2.8: Experiment I: The effect of home ownership on labor market outcomes

	Simulated Renters	Simulated Owners	Difference
Unemployment rate	1.55 %	1.38%	-0.17%
Monthly wage	3,545	3,512	-33
Unemployment duration(weeks)	14.66	13.34	-1.32
End with local jobs(weeks)	13.88	12.32	-1.56
End with non-local jobs(weeks)	17.21	18.98	1.77

Table 2.9: Experiment I: The effect of home ownership on labor market outcomes (different demographic groups)

	Difference in unemployment durations (weeks)
low education, ages 20-39;	-1.97
high education, ages 20-39;	-1.25
low education, ages 40-55;	-1.01
high education, ages 40-55;	-0.24

Figure 2.4: Model fit: Unemployment rate and wages by age

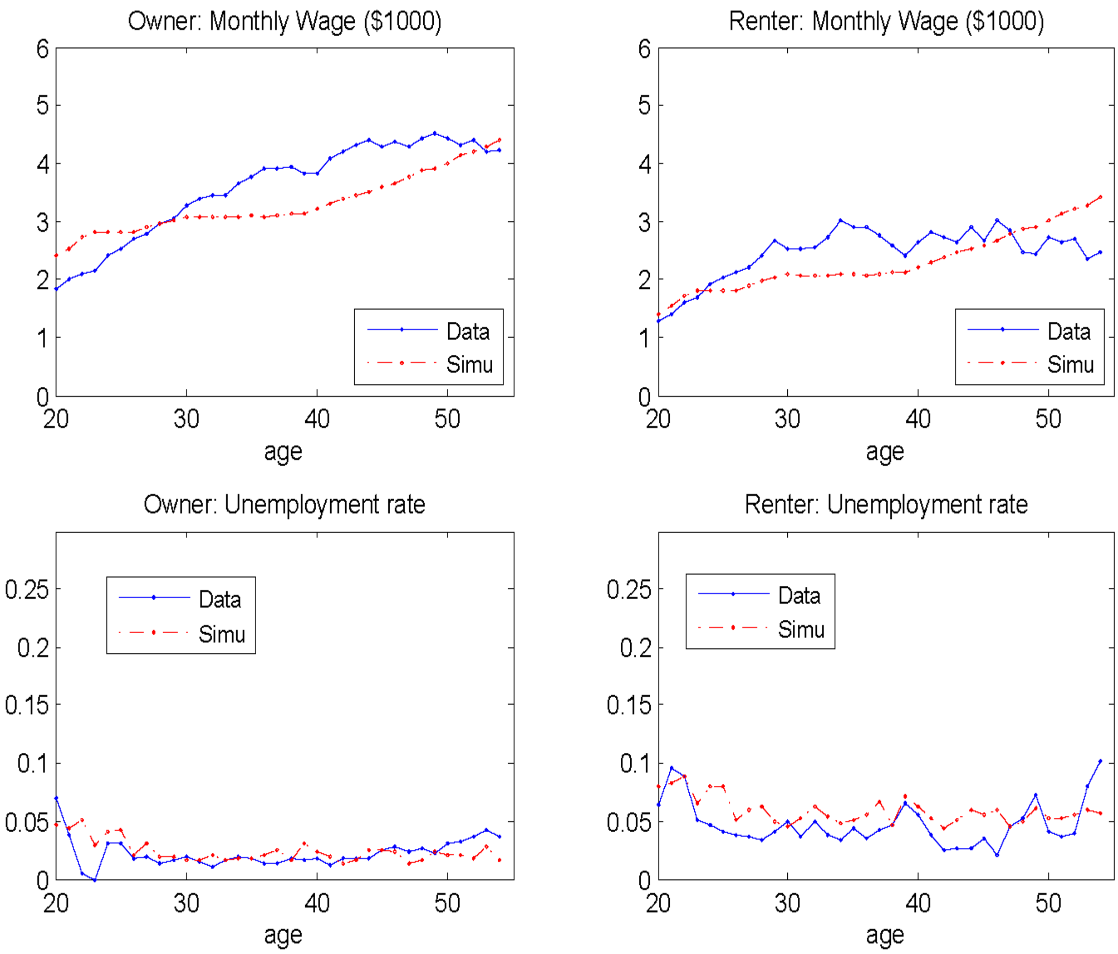


Table 2.10: Experiment II: Alternative moving cost (ξ^m)

	Baseline $\xi^m = 797$	Experiment $\xi^m = 396$	Experiment $\xi^m = 0$
Unemployment rate	-0.17%	-0.08%	-0.03%
Monthly wage	-33	-19	-12
Unemployment duration	-1.32	-0.53	-0.04
End with local jobs(weeks)	-1.56	-0.64	-0.09
End with non-local jobs(weeks)	1.77	0.69	0.04

Table 2.11: Experiment III: Alternative labor market environment

	Baseline	Experiment
No job offer from non-local market ($\lambda_l = 0.9, \lambda_n = 0$)		
Unemployment rate	-0.17%	-0.15%
Monthly wage	-33	-27
Unemployment duration(weeks)	-1.32	-0.53
End with local jobs(weeks)	-1.56	-0.53
End with non-local jobs(weeks)	1.77	0
No job offer from local market ($\lambda_l = 0, \lambda_n = 0.9$)		
Unemployment rate	-0.17%	0.11%
Monthly wage	-33	10
Unemployment duration(weeks)	-1.32	1.59
End with local jobs(weeks)	-1.56	0
End with non-local jobs(weeks)	1.77	1.59

Table 2.12: Experiment IV: Eliminate mortgage interest deduction in alternative labor market environment

	Baseline	Experiment	Difference
Eliminate Mortgage Interest Deduction ($\lambda_l = 0.8, \lambda_n = 0.1$)			
Home ownership rate	69%	55%	-14%
Unemployment rate	1.45 %	1.47 %	0.02 %
Monthly wage	3,558	3,542	-16
Unemployment duration (weeks)	14.66	15.05	0.39
End with local jobs(weeks)	13.87	14.31	0.44
End with non-local jobs(weeks)	17.21	17.02	-0.19
Eliminate Mortgage Interest Deduction ($\lambda_l = 0.45, \lambda_n = 0.45$)			
Home ownership rate	69%	52%	-17%
Unemployment rate	1.45 %	1.42 %	-0.03%
Monthly wage	3,558	3,536	-22
Unemployment duration (weeks)	14.66	14.18	-0.48
End with local jobs(weeks)	13.87	13.34	-0.53
End with non-local jobs(weeks)	17.21	17.63	0.42

Chapter 3

Are Women Working More to Pay the Mortgage? Evidence from SIPP 1996

3.1 Introduction

Purchasing a house is one of the most important financial decisions a family makes. Living in an owned house usually entails a long-term financial commitment, which affects not only how family members live, but also, not surprisingly, how they work. Home ownership affects people's labor market outcomes in two ways. On the one hand, compared with renters, homeowners are more closely tied to the local market and are less likely to move for job opportunities when unemployed. A large literature studies the relationship between home ownership and unemployment and is still debating about the signs and the size of this mobility effect[1]

On the other hand, the mortgage payment associated with home ownership encourages homeowners, especially wives, to participate more in the labor market. This positive effect could be derived from two characteristics of housing consumption. First, like the other types of consumption, housing consumption provides positive utility. To enjoy more housing consumption, people have an incentive to work more. As in the context of the labor supply model, this is the case in which the marginal

[1] See Oswald (1996) Green and Hendershott (2001),Coulson and Fisher (2002),Coulson and Fisher (2009),Flatau, Forbes, and Hendershott (2003), Munch, Rosholm, and Svarer (2006),Battu, Ma, and Phimister (2008), and Van Leuvensteijn and Koning (2004).

disutility of working is smaller than the marginal utility of housing consumption. Second, unlike non-durable good consumption, housing consumption can be viewed as a commitment expenditure[2] because it usually is subject to a long-term mortgage contract. According to the contract, a substantial amount of household total income has to be used for housing consumption each period, and the adjustment of this consumption is subject to a fixed cost. This feature of housing consumption reduces households' ability to buffer the uncertainty about their future income(Carroll (1997) and Carroll and Kimball (2001)). For example, when there is a negative income shock, it's harder for households with mortgages to smooth consumption in response to this shock. Instead, they are more likely to adjust their labor market behaviors. Therefore, households with mortgages tend to work more today to build up a buffer stock of assets in case of a future negative income shock. This is especially true for females, whose labor market choices are more flexible.

To summarize, there are two mortgage related driving factors for the female labor supply, corresponding to the two characteristics of housing consumption. The first is active, which arguing that females would want to work in order to enjoy more housing consumption. The second is passive, arguing mortgage payments reduce the ability to buffer future risk through consumption and forces female to work more in order to build up a buffer stock of asset. The second mortgage related drive is called the commitment effect. Contrary to the mobility effect, which discourages people from working, the commitment effect encourages wives to enter the labor force.

This paper tries to identify the positive side of home ownership by investigating how the size of the home mortgage impacts the female labor supply. It examines the relationship between female labor supply and household mortgage holding using data from SIPP 1996 Panel. The estimation sample consists of married women who

[2] The same idea is proposed and modeled in Chetty and Szeidl (2010).

are between 20 and 65 years old.

Most of the previous literature concerning the relationship between mortgage-related variables and the female labor supply has usually failed to control for wealth characteristics when conducting an estimation. The SIPP data contain a detailed set of questions about assets, debts, income and financial characteristics of the household that allows for a more complete analysis of the relationship between the two variables.

An important issue in this research is the potential endogeneity of mortgage status. Endogeneity may appear due to the simultaneity between mortgage commitments and the female labor supply or due to reverse causality. Simultaneity may arise if, for example, there are unobserved characteristics that affect both the labor market decision and the mortgage decision. The reverse causality emerges if the labor supply affects the mortgage decision. For instance, banks take into account the employment status of the wife when the household takes out a mortgage. To address this endoengeity problem, I use the state-level mortgage interest deduction as an instrumental variable. This tax policy affects home mortgage decisions, (Hilber and Turner (2010)) but has no direct correlation with labor market outcomes.

The estimation is based on the standard static female labor supply model, where selection into the labor force(unobserved wage rate for non-labor market participants) is solved by the Heckman two-step method(Heckman (1979)). The estimation results indicate that wives in households that hold mortgages are more likely to participate in the labor market. Different specifications come to the same qualitative results. I also estimate the regression model for subsamples of couples with large and small household wealth, finding that wives with limited household wealth are more likely to be affected by household mortgage status.

The organization of the paper is as follows. Section 2 presents a review of the related literature. Section 3 presents the theoretical framework used to predict the effect of mortgages and guide the empirical specification. Section 4 introduces the data and provides a statistical description of the main variables. Section 5 presents the empirical strategy and reports the estimation results for the baseline model. Section 6 conducts robustness checks. Section 7 concludes.

3.2 Literature review

A number of papers investigate the impact of housing financial status on household labor supply based on micro-data in different countries. Yoshikawa and Ohtaka (1989) examines household savings, the labor supply of married women, and the demand for residential housing in Japan, considering a three-period life-cycle model of household behavior. Their results suggest that the labor supply of married women in Japanese homeowner households is an increasing function of the price of land. Fortin (1995) analyzes the effect of mortgage qualification constraints on patterns of female labor supply in Canada. She uses a cross-section data to estimate a life-cycle model of household labor supply that incorporates a mortgage qualification constraint based on earnings. She finds that a nontrivial percentage of married women are constrained by mortgage commitments. Bottazzi (2004) uses panel data from the British Household Panel Study between 1993 and 2000 to analyze whether mortgage commitments have any effect on female labor market participation. Specifically, she considers a life-cycle model in which a mortgage qualification constraint imposed by banks holds at each time period, as in Fortin (1995). The endogeneity problem is addressed by using housing prices as an instrument for the obligation ratio (ratio between monthly mortgage payment and household income exclusive of the female's labor income). It is found that the obligation ratio has a positive effect on female participation and that the endogeneity of the obligation ratio cannot be rejected.

This paper is related to the literature on housing commitment consumption. Chetty and Szeidl (2010) analyze the effect of commitment consumption, including housing consumption, on households' risk aversion. They find that households are more risk-averse when committed to certain consumption. Several recent studies also explore the implications of commitment consumption in other contexts, including Flavin and Nakagawa (2008), Postlewaite, Samuelson, and Silverman (2004), and Shore and Sinai (2010). None of these studies explores the implications of commitment consumption for labor market outcomes.

Last but not least, this paper contributes to the literature on the second generation of female labor supply, which emphasizes the the sample selection bias caused by the missing wage for non-workers(Killingsworth and Heckman (1986),Blundell and MaCurdy (1999)). This paper solves this sample selection problem with traditional the Heckman two-step method(**?**). More importantly, this paper considers the impact of housing financial status on female labor supply and studies whether the estimation of the labor supply equation changes when taking the housing variable into consideration. This gives a more comprehensive picture of female labor supply.

3.3 Theoretical Framework

To illustrate how housing consumption affects the wife's labor market decisions, I consider a simple static labor supply model in which wife makes decisions among leisure(L), housing consumption (H)[3], and non-housing consumption(C). The current utility function is then denoted as $U(C, H, L)$, and the wife's maximization

[3]Depending on the tenure status, housing consumption is paid as rent or a mortgage. This paper focuses on the effect of mortgage payments on labor supply and excludes the effect of tenure choices. Therefore, I consider only the households that are living in an owned house.

problem is expressed as follows:

$$\max U(C, H, L)$$

s.t.

$$N + W(T - L) = C + PH,$$

where N is exogenous non-labor income[4];W is wage; P is the normalized price of housing consumption on the price of other consumption; and T is the wife's total time endowment. Then, WL is the total labor income of the wife, and $N+WT$ is the full income that she can use to purchase consumption and leisure. $U(\cdot)$ are assumed to be strictly concave and monotonically increasing in C , H and L, respectively. Let λ be the Lagrange multiplier associated with the budget constraints. We have the following first-order conditions with respect to L,C,and H[5]:

$$U_c(C, H, L) = \lambda \tag{3.1}$$

$$U_h(C, H, L) = \lambda P \tag{3.2}$$

$$U_l(C, H, L) = \lambda W \tag{3.3}$$

To get ride of the Lagrange multiplier and put the first-order condition in an alternative form, we have

$$\frac{U_l(C, H, L)}{U_c(C, H, L)} = W \tag{3.4}$$

$$\frac{U_l(C, H, L)}{U_h(C, H, L)} = \frac{W}{P} \tag{3.5}$$

$$\frac{U_h(C, H, L)}{U_c(C, H, L)} = P \tag{3.6}$$

[4]Non-labor income includes husband's labor income, which is taken as exogenous here because men typically work full-time and, therefore, their labor supply is less flexible.

[5]Let's leave corner solutions alone for now.

Equation (4) is the same first-order condition as in the classical static labor supply model[6], while equations (5) and (6) describe how the price of housing consumption (P) affects the wife's labor supply as well as non-durable consumption.

Housing consumption and labor supply are related in two ways. First, like the other types of consumption, housing consumption provides positive utility. Wanting more housing service gives people the incentive to work more. This is the case when the marginal utility of working is smaller than the marginal utility of housing consumption. More specifically, suppose that P goes down, as happens when the mortgage deduction goes up; the cost of housing consumption is lower, and, thus, the housing consumption will go up according to equation (2). At the same time, decreasing in P also has income and substitute effects on non-durable consumption C and labor supply $T - L$. I pay special attention to the wife's labor supply. Due to the income effect, wives are able to afford more consumption with the same amount of labor supply and, thus, may tend to work less; due to the substitution effect, in order to enjoy greater housing consumption, the wife wants to work more. The net effect depends on whether the income effect dominates the substitution effect, or the other way around. This paper tests which effect is dominant by looking at the relationship between housing consumption and the wife's labor supply.

Second, unlike non-durable good consumption, housing consumption can be viewed as a commitment expenditure because it is usually subject to a long-term mortgage contract. According to the contract, a substantial amount of household total income has to be used for housing consumption in every period, and the adjustment of this consumption comes along with a fixed cost. Facing the uncertainty in future income, those with a mortgage can not react as flexibly as those without a mortgage. For example, when there is a negative shock in the household income,

[6] See Blundell and MaCurdy (1999).

the household might want to reduce their housing consumption. However, due to the adjustment cost, reducing housing consumption might not be optimal, compared with keeping the amount of housing consumption as the same in the last period. In other words, the adjustment cost causes a utility loss. To compensate this loss and enjoy the same level of life-time utility, wives might need to work more when the disutility of working is small. In other words, a mortgage commitment reduces the household's flexibility of allocating income to housing and non-durable consumption because the disposable income is smaller due to the adjustment cost.

The existence of the adjustment cost for housing consumption can alter people's optimal labor supply. This is a commitment effect and we can illustrate it in a simple two-period model.

$$\max U(C_1, H_1, L_1) + \beta EU(C_2, H_2, L_2) \tag{3.7}$$

s.t.

$$N_1 + EN_2 + W_1(T - L_1) + EW_2(T - L_2) = C_1 + C_2 + PH_1 + PH_2, \tag{3.8}$$

which is an intertemporal budget constraint. N_1 and W_1 are the non-labor income and wage for period 1, while N_2 and W_2 are the non-labor income and wage for period 2. When the agents make optimal consumption and labor supply decisions, they know current but not future income. They know the distribution of future income and makes decisions based on this information. The uncertainty about future income induces a precautionary motive. As illustrated in Carroll (1997) and Carroll and Kimball (2001), due to the lack of completeness of insurance markets, the precautionary motive decreases current consumption in order to smooth current and future consumption. However, housing consumption is relative rigid, which, in turn, reduces households' ability to buffer the risk in future income. In response to this rigidity in consumption, households would adjust their labor market decisions–that is, wives would work more and bring in additional labor income to buffer future risk.

The key element of this story is the rigidity of housing consumption. To illustrate this point, I assume that the adjustment of housing consumption is subject to a cost kH,[7]where k is a positive parameter. Since large houses usually induce large transaction costs, it's reasonable to assume this adjustment cost is larger when the current housing consumption is large. For simplicity, I assume away the interest rate.

The optimal solution of this two-period model has two cases, depending on whether the agent changes the housing consumption between the two periods. If the agent keep the housing consumption the same for the two periods, the optimization problem can be viewed as the same as adding an extra constraint

$$H_1 = H_2 \tag{3.9}$$

to the above optimization problem; if the agent changes the housing consumption between the two periods, she faces a new intertemporal budget constraint

$$N_1 + EN_2 + W_1(T - L_1) + EW_2(T - L_2) - kH_1 = C_1 + C_2 + PH_1 + PH_2, \tag{3.10}$$

where the adjustment cost kH_1 is imposed. Let's call the optimization problem of the two cases $P1$ and $P2$, and the original optimal problem without any obligation of housing consumption $P0$. It's easy to see that the optimal two-period utility is no larger in $P1$ and $P2$ than that in $P0$, because the feasible sets of $P1$ and $P2$ are smaller than that in $P0$.

The existence of the adjustment cost of housing consumption can alter households' optimization problem and, therefore, alter their labor supply decisions. In other words, the optimal choices of labor supply in $P1$ and $P2$ are different

[7]It does not matter whether k is a real monetary cost or a psychological cost incurred due to moving.

from those in $P0$. I leave the theoretical proof of how this commitment effect directly impacts labor supply for future research.[8] But I emphasize the intuition of the commitment effect in this paper. Housing commitment or, to be more precise, the adjustment cost associated with housing consumption reduces the flexibility of household income allocation and the ability to buffer the uncertainty about the future. Wives then tend to work more to increase the disposable income and to better deal with future risk. The larger the adjustment cost–that is, the mortgage obligation–the larger the commitment effect. I test this hypothesis in the empirical section.

Housing consumption, usually is financed by a long-term mortgage contract, which requires certain amount of expense every period and the adjustment requires a fixed cost. This means the housing consumption is relative rigid than other type of consumption. Having a mortgage distorts wives' labor market decision and make them work more in order to enjoy more housing service and pay off the mortgage debt. Therefore, I propose the hypothesis that the income effect is small and is dominated by the substitute effect, which means increasing in housing consumption (mortgage payment) would increase wife's labor supply. The empirical work in the following sections are for the purpose of test this hypothesis.

3.4 Data

3.4.1 SIPP 1996 Panel

The Survey of Income and Program Participation (SIPP) 1996 panel covers 1996-1999 and collects data every four months (12 periods in total). Each period provides comprehensive information on demographic characteristics and labor force status, indicating whether the agent is working, looking for a job or is out of the labor

[8] Chetty and Szeidl (2010) and Chetty (2006) illustrate the commitment effect in a dynamic model, and their theoretical results suggest that commitments may induce important changes in household risk aversion and consumption decisions.

force in current month.[9] In addition, SIPP contains detailed information about the financial situation of the household, different sources of labor and non-labor income, the value of real and financial assets and the value of mortgage and personal debts. In particular, the household is asked about the type of housing tenure, the existence of an outstanding mortgage, the year in which they bought the house, and the initial and current values of the property. This information allows for a more complete analysis of the relationship between the housing financial situation and labor market outcomes within a household. Another advantage of SIPP is its relatively large sample size. The original SIPP 1996 panel has 3,658,293 person-month observations.

The sample used in the empirical work is selected from the original sample that contains only married females between 20 and 65 years old. Since this paper focus on mortgage payment, tenants and other types of households are excluded from the. Since home ownership and labor force status are the main variables in this paper, I also drop observations whose residential status or mortgage status are missing, as well as those for individuals enrolled in school or the Armed Forces. The final sample is an unbalanced panel[10] containing 57,318 observations. All monetary variables are converted to 1996 dollars.

3.4.2 Variables

This paper considers three labor market variables: (1) Labor participation: SIPP contains data on monthly employment status(employed, unemployed and out of labor force):[11] The individual is defined as a participate in the labor market if

[9]Although SIPP provides monthly information, data for the first three months of each wave are collected from retrospective questions that might be less reliable than these for the interview month. To avoid the unnecessary seam bias, I choose a wave (four mouths) as the time unit.

[10]This sample reduction occurred due to the normal survey attrition, such as refusing to continue to participate, inability to locate persons, deaths, etc.

[11]In SIPP, the basic labor force information has been recoded into eight employment status recodes (ESR's). These ESRs are defined as follows:
ESR 1 –With job entire month, worked all weeks.
ESR 2 –With job entire month, missed 1 or more weeks, but not because of a layoff.

she is employed or unemployed and actively looking for a job at certain time of the month[12]. Additional definitions of female participation will be considered to check the sensitivity of the results;

(2)Hours of work: For each month, respondents report hours of work per week and how many weeks worked. Monthly labor supply is calculated as hours per week(weeks worked/weeks in month)$\times$ 4.33;

(3)Hourly wage: For those who are not paid by the hour, their hourly wages are obtained by dividing monthly earning by the hours of work provided during that month. For those who are paid hourly, their hourly wages are measured by their hourly rate. Wage is observed only for those who are participating in the labor market.

SIPP asks questions on the ownership status of their living quarters. Thus, homeowners are defined as those who own the living quarters. Renters are defined as those who are paying rent, excluding those who live in public housing. To pay special attention to the effects of a mortgage, I define the mortgagor as a dummy variable that takes the value one if the household declares any payments outstanding on a mortgage taken out to buy the main residence and zero otherwise. To quantify the mortgage status, I measure the mortgage as the total debt owed on the current house.[13] The basic idea is that a larger mortgage is associated with larger financial pressure, which forces wives to work more. To further measure the mortgage commitment, I use the ratio of mortgage payment to family income.

ESR 3 –With job entire month, missed 1 or more weeks because of a layoff.
ESR 4 –With job part of month, but not because of a layoff or looking for work.
ESR 5 –With job part of month, some time spent on layoff or looking for work.
ESR 6 –No job in month, spent entire month on layoff or looking for work.
ESR 7 –No job in month, spent part of month on layoff or looking for work.
ESR 8 –No job in month, no time spent on layoff or looking for work.

[12]I define those whose ESR=8 as non-participating.

[13]This measurement is different from the debt-to-income ratio, which is defined as total minimum monthly debt (home mortgage included) divided by gross monthly income.

In order to control for income and wealth effects in the labor supply decision, I include the spouse's income, other household income(wife's labor income is excluded to avoid endogenous problem) and household wealth in the analysis. Other demographic variables used in the analysis include the age of the wife, the number of young children (0-6) and the number of old children (7-15), whether the household lives in a metropolitan area, year dummies, and three dummy variables indicating the wife's education level. The educational levels considered are high school education, college education, and graduate education. Those with a college education include those who graduated from high school and may have attended college but not received a four-year college degree.

3.4.3 Descriptive Statistics

Figure 1 describes female participation rates and hours of work by different mortgage statuses. The key feature that emerges from the figure is that the existence of a mortgage is associated with a higher labor participate rate and more hours of work. On average, wives who are currently holding mortgages have a labor participate rate of 78% and work about 124 hours every month, while only 62% of wives without a mortgage participate in the labor market and work about 98 hours every week. This is consistent with the hypothesis that wives who need to pay mortgage are active in the labor market. However, females with no mortgage payment tend to be older, and at the same time, more likely to be out of the labor market. To look at the relationship between labor supply and mortgage status, I divide wives with mortgages into three groups by the size of their mortgage.[14] Figure 1 shows that a larger mortgage is roughly associated with longer working hours and a higher labor participation rate, though the difference is a little smaller than that between those who hold mortgages

[14]The three groups are wives with a mortgage below the 50th percentile, between the 50th and the 75th percentile, and above 75 percentile. The sample size of each group of females is roughly the same.

and those who do not. For example, for wives between ages 20 and 35, those with relatively small mortgage(below \$50,000) have a lower labor participate rate of 77% and work about 123 hours every month, while those with a larger size of mortgage (between \$ 50,000 and \$ 90,000) have a higher labor participate rate of 80% and work about 128 hours every month. As the figure shows, those features are captured within different age groups. The difference between mortgagors and non-mortgagors for wives above 50 years of age is the largest among the three age groups, indicating that mortgage status might have a large impact on female's retirement decisions.

Summary statistics of the variables used in the estimation are presented in Table (3.1)(in 1996 dollars for monetary variables). During the survey period, about 76.4% married women participated in the labor market, and they worked about 121 hours per month. About 74.5% married women lived in a house with a mortgage, and the average mortgage ratio was 12.1%, which is measured by the ratio of mortgage payment to family income. About 48% married women had college or higher education. The hourly wages of the wife and spouse were 12.5\$ and 16.7\$. We can observed wages only for those who were working, which is missing for more than a tenth of the sample. The monthly spousal income and household income were \$2,718 and \$4,621, while the average household wealth was \$116,955.

3.4.4 Mortgage Tax Subsidies

Federal and state income tax policies affect the cost of home ownership and mortgage status. This paper focuses on the mortgage interest deduction, which is the most important favorable tax treatment for home owners. The basic idea is that when the deduction is higher, homeowners have a strong incentive to hold a larger mortgage. To measure this deduction, I calculate the tax savings from an additional dollar of mortgage interest, the mortgage interest subsidy, based on NEBR's publicly available data on tax rates.[15] To be more specific, the mortgage interest subsidy rate is calculated as a tax saving from an additional dollar of mortgage interest. I first calculate the state income tax liabilities owed by a representative sample of taxpayers in SIPP.[16] Then, I increase the mortgage interest by 1% for the taxpayer and recalculate the state taxes. Then, the mortgage interest subsidy rate is generated as the ratio of the additional tax (savings) to the additional 1% mortgage interest.

The average mortgage interest subsidy for different years and different s-tates in the period 1996-2000 [17] is presented in Table (3.2) and Table (3.3). There are large differences in this tax subsidy across different states.[18] [19]Some states, such as Florida, Nevada and Texas, collect no personal income tax at all, while others, such as California, Delaware, Maine, Massachusetts and North Carolina, rely heavily on personal income taxes to raise revenue, but permit the deduction of mortgage interest. Among these states, the mortgage subsidy rate varies considerably, reaching a maximum of around 9% per dollar of mortgage interest in the District of Columbia. The correlation between the state-level mortgage interest deduction and other state-level variables, including annual income per capita, unemployment rate and housing prices, is not significant. The comparison of states with and without the mortgage interest deduction also shows no significant difference in income, household characteristics or labor market outcomes. In addition, the deduction variation across states

---

[15]See details in Feenberg and Coutts (1993) and at http://www.nber.org/taxsim.

[16]Representative taxpayers with high, median and low annual income has been tried, and the main estimation results are the same.

[17]I cannot identify all 50 states because in the SIPP 1996 panel, Maine and Vermont, as well as North Dakota and South Dakota share the same state code.

[18]Current mortgage interest subsidy captures the state-level variation of the tax policy, by calculating the mortgage interest subsidy based on individual level characteristics, more variation will be allowed in the future work.

[19]One concern of using state-level instrument is that some people might work in different state than they live in. Since not many people are living along the border of states and have the options to live in either states, these sample would not have large influence on the regression results.

is usually determined by federal and state tax laws.[20] These facts suggest that the mortgage interest deduction is exogenous to individuals' labor market decisions.

3.5 Estimation

3.5.1 Estimation Model

The empirical model in this paper is built on the second-generation labor supply studies[21], in which labor supply is given by the combination of the participation decision H_i and hours of work H_i^*. As in the static labor supply model illustrated in the last section, the wife's labor supply is determined by wage and the marginal rate of substitution of goods for leisure and housing consumption. This relationship can be expressed by the following empirical model:

$$H_{it}^* = \beta_0 + \beta_1 M_{it} + \beta_2 W_{it} + \beta_3 X_{1it} + \epsilon_{it}, \tag{3.11}$$

$$H_{it} = \begin{cases} 1; & \text{if} \quad H_{it}^* > 0 \\ \\ 0; & \text{if} \quad H_{it}^* = 0, \end{cases} \tag{3.12}$$

or

$$H_{it}^* = \beta_0 + \beta_1 M_{it} + \beta_2 W_{it} + \beta_3 X_{1it} + \epsilon_{it}, \tag{3.13}$$

$$H_{it} = 1\{\beta_0 + \beta_1 M_{it} + \beta_2 W_{it} + \beta_3 X_{1it} + \epsilon_{it} > 0\}, \tag{3.14}$$

where M_{it} is the mortgage ratio measuring the mortgage status, W_{it} is wage, and X_{1it} are other controls that are correlated with labor supply decisions, such as the number of children, and household income. Spousal income is taken as exogenous as part of other household income because, typically, men are already working full-time and, therefore, their labor supply is less flexible. I pay special attention to β_1, which measures the relationship between labor supply and mortgage status.

---

[20]Different states implement different formulas for taxable income; some use federal adjusted gross income as a starting point for developing their tax base, while others use federal taxable income. And the taxable income in other states is computed independently of the federal formula.

[21]See Killingsworth and Heckman (1986) for a survey.

There are two problems with estimating the causal parameter β_1 directly from equation (11). The first one is the well-known sample selection problem, as illustrated in Heckman (1979). I observe only the wages of wives who participate in the labor market ($H_i > 0$). If I estimate the above equation using only data on workers with $H_i > 0$, the estimates of the β parameters will suffer from sample selection bias, which usually results a downward bias. On the other hand, if I include non-workers $H_i = 0$, this surely gets the nonlinearities wrong. To get around this problem, as suggested in Heckman (1979), I need to jointly model wages and participation. The wage equation is expressed as follows:

$$W_{it} = \alpha_1 Z_{1it} + \mu_{it} \quad \text{if} \quad H_{it} > 0, \tag{3.15}$$

where Z_1 is a vector of observed characteristics that affect wage, and μ_{it} is unobserved characteristics. Education affects people's skill and wage. It may also affect tastes for working, but that effect is likely to be smaller than the direct wage effect. Another variable is whether the wife lives in a metro area. Metro areas are usually associated with higher wages than non-metro areas because of the high demand for labor. Even though the tastes for work may be different in metro and non-metro areas, this effect is likely to be smaller than the wage effect. So I include education and whether the wife lives in a metro area in the wage equation Z_1, but not in the labor supply equation X_1.

The second problem of directly estimating equation (11) is the endogeneity of the mortgage status variable M_{it}. Endogeneity may appear due to simultaneity between mortgage commitments and the female labor supply or due to reverse causality. On the one hand, some unobserved characteristics might be correlated with both the labor supply choice and the mortgage choice. For example, there might be a group of people who play hard and work hard. It's not the large mortgage that forces them to work more; simply enjoy having a large house and prefer to work hard at the same time. However, I could not observe these preferences from the data. On the other hand, reverse causality emerges if labor supply affects the mortgage decision. For instance, bank takes into account the employment status of the wife when the household takes out a mortgage.

Suppose that mortgage status M_{it} is determined by the following equation:

$$M_{it} = \alpha_2 Z_{2it} + \eta_{it} \tag{3.16}$$

where Z_2 is a vector of observed characteristics that affect mortgage choices, and η_{it} is unobserved characteristics. If η_{it} is correlated with the unobserved error ϵ in the labor supply equation, then M_{it} is endogenous, and the estimation without taking this problem into consideration is biased. To solve this problem, I include the state-level mortgage subsidy variable in Z_2 to provide identification.

To summarize, the empirical model consists of the following equations:

$$H^*_{it} = \beta_0 + \beta_1 M_{it} + \beta_2 W_{it} + \beta_3 X1_{it} + \epsilon_{it}, \tag{3.17}$$

$$H_{it} = \begin{cases} 1; & \text{if} \quad H^*_{it} > 0 \\ 0; & \text{if} \quad H^*_{it} = 0, \end{cases} \tag{3.18}$$

$$W_{it} = \alpha_1 Z_{1it} + \mu_{it} \quad \text{if} \quad H^*_{it} > 0 \tag{3.19}$$

$$M_{it} = \alpha_2 Z_{2it} + \eta_{it} \tag{3.20}$$

3.5.2 Estimation Method and Estimation Results

The empirical strategy in this paper can be viewed as an adaptation of the selection-corrected methodology in Heckman (1979), with an extra step added to deal with the endogeneity problem. It can also be viewed as a two-stage least-squares method with an inverse Mills ratio to solve for the sample selection problem[22]. The

[22] See Semykina and Wooldridge (2010) and Wooldridge (2010) for more detail on this method.

estimation consists of three steps. In the first step, as in Heckman (1979), I estimate the wage equation, controlling for the selection in the equation with the inverse Mills ratio. Then, I predict wages for all women based on this wage equation. In the second step, I estimate the linear mortgage equation with the instrumental variable–mortgage interests subsidy–and obtain a predicted mortgage status. In the third step, the labor supply equation is estimated with the predicted wage and predicted mortgage status.

An alternative estimation strategy for the above empirical model is the maximum likelihood method, which typically has a substantial efficiency advantage. I choose the three-step method for two reasons. First, it is easy to implement and numerically robust because it doesn't require one to numerically maximize a complicated likelihood function. More importantly, this approach can relax the strong normality assumption.

In the first step, since I do not observe the wages for the women that do not work, I impute the non-observed wages using the traditional regression imputation method.[23] The wage is imputed based on a set of individual characteristics, including Z_1 (constant, age, age-square, race, education dummies, metro variable) and the inverse Mills ratio computed to correct for the endogenous labor force participation. The wage equation is

$$\hat{W_{it}} = \alpha_1 Z_{1it} + \rho\hat{\lambda} + \mu_{it}^*, \text{if} \quad W_i > 0, \tag{3.21}$$

where the inverse Mills ratio is calculated by

$$\hat{\lambda_{it}} = \frac{\Phi(\hat{\beta} X1_{it})}{\phi(\hat{\beta} X1_{it})}, \tag{3.22}$$

where $X1$ includes variables that explain a person's participation in the labor market (constant, age, age-square, race, number of young and old children, household income

[23]Some authors, e.g. Morissette and Hou (2008)), use a quantile wage of the employed women.

net of wife's labor income, education and metro variable). The method up to this point is simply the Heckman two-step method. The exclusion of household income and children variables in this wage equation provides identification of the inverse Mills ratio term other than what would come from functional form assumption alone. Wages for all women are predicted base on the wage equation using the estimation of α_1. Table (3.4) reports the results for the first step. Column one reports the results of the probit participation regression, which is used to calculate the inverse Mills ratio. Column two reports the regression results of the wage equation with the inverse Mills ratio.

In the second step, I estimate the following mortgage status equation,

$$M_{it} = \alpha_2 Z_{2it} + \eta_{it}, \tag{3.23}$$

where M_{it} is the mortgage ratio measured as the ratio of mortgage payment to household income, and Z_2 is a set of variables affect mortgage status, including state-level interest deduction, age, race and education dummies, number of children, whether the household lives in a metro area and household income. Table 5 presents the regression results of the mortgage status equation. The first column presents results for pooled OLS, while the second and third columns present results for fixed-effect OLS and random-effect OLS, controlling unobserved heterogeneity across individuals. As presented in Table (3.5), the mortgage interest subsidy is significantly correlated with the mortgage ratio in all three models, suggesting that people are more likely to hold large mortgage debt in the states in which tax policies are more favorable to owners.

In addition, other households income and number of children are also positively related with mortgage ratio. Those with higher other monthly income and those with more young or old children in the households are more likely to hold larger mortgages. The Hausman test shows that the p-value is 0[24] and that the fixed-effect model is more appropriate than the random-effect. Therefore, I predict mortgage status based on the fixed-effect model.

In the third stage, I consider the following selection-corrected regression model for labor supply. First, I consider a linear regression model for hours of wage H^*_{it},

$$H^*_{it} = \beta^h_0 + \beta^h_1 \hat{M}_{it} + \beta^h_2 \hat{W}_{it} + \beta^h_3 X1_{it} + \eta^h_{it}, \tag{3.24}$$

where H^*_{it} is the number of hours of work per month; $\hat{W}_{it}$ is the predicted wage from the first step; and $\hat{M}_{it}$ is the predicted mortgage status from the second step. The instrumental variable is not correlated with the error term in the labor participation equation; that is, $E(\eta^h_{it}|Z_{it}) = 0$. Then, the causal effect is measured by β^h_1, which is the focus of the research. $X1_{it}$ is the set of variables correlated with labor supply decisions, including age, race, number of young and old children, household income,[25] and year dummies.

Second, I consider a probit model[26] for the labor participation variable H_{it}:

$$H_{it} = \beta^p_0 + \beta^p_1 \hat{M}_{it} + \beta^p_2 \hat{W}_{it} + \beta^p_3 X1_{it} + \eta^p_{it}, \tag{3.25}$$

where H_{it} is a binary indicator with 1 indicating that the agent participates in the labor market. Other variables are the same in the hours of work equation.

---

[24] chi2=89.43

[25] Another common form of the labor supply model would include the spouse's wage to allow analysis of substitution or complementarity of husband's and wife's labor supply. This is not the concern in this paper, and I don't differentiate the spouse's income from other household income.

[26] Limited dependent variables call for nonlinear models like probit. These nonlinear models face special challenges when there are endogenous regressors. One method is merely to estimate a linear probability model using IV. As Wooldridge (2002, p.472) says, " This procedure is relatively straightforward and might provide a good estimate of the average effect." Also see Angrist (2001) and Angrist, Imbens, and Rubin (1996) about the justification of the implications of the linear model estimation techniques in such cases. Another alternative estimation method of the probit model with endogenous regressors is provided by Rivers and Vuong (1988). Despite the fact that these estimators rely on strong distributional assumptions, they are implemented in standard econometric software packages (such as STATA) and are still frequently used in applied work.

Table (3.6) summarizes the regression results of the baseline model. The first column corresponds to the estimation of hours of work equation. The second column corresponds to the estimation of the probit model for the labor participation. The causal effect parameter, coefficient for the mortgage ratio is significantly positive in both models, which show the mortgage debt has positive effects on the labor supply of married women. On average, one percent higher in mortgage ratio encourage women to work about 0.55 hour more every month. It also increases the probability of participating in the labor market by about 0.3%. The empirical evidence presented above clearly shows that mortgage commitments influence the labor market decisions of married women. The implication here is that the maximum gross debt service ratio allowed by the lending institutions may have a significant impact on the labor supply of married women. Further, since mortgage status and housing policies prevent wives from exiting the labor market at will, they could have a significant effect on family life and childbirth.

The wage elasticity is estimated to be about 2.89, which lies in the range of wage elasticity estimated in the literature. Both the number of young children (0-6) and number of old children (7-15) are significantly negative in explaining the female's hours of work and the participation propensity. Married women with children usually work fewer hours and are less likely to participate in the labor market. The coefficient for household income net of the wife's labor income is negative and significant across different models, which implies a negative income effect on labor supply. These results are consistent with those in the literature.[27]

Lastly, the spouse's income plays an important role in the labor supply of wives. To look at how spouse's income affects the effects of mortgage status, I decompose the household income net of female labor income into two terms, the spouse's

[27] See Killingsworth and Heckman (1986) and Blundell and MaCurdy (1999).

income and other household income, where other household income is the household income subtracted from the wife's and her spouse's income. Table (3.7) summarizes the regression results for the new model specification. The first column corresponds to the estimation of hours of work equation. The second column corresponds to the estimation of the probit model for the labor participation. As in the baseline model, married females with larger mortgage ratio work more and have a higher probability of participating in the labor market. When the household income is decomposed, the coefficient for spouse's income is large and negative, much larger than the coefficient for other household income. This income effect comes mainly from spouse's income.

3.6 Robustness check

In this section, I check the robustness of the baseline results by conducting the estimation for different models and different subsamples.

3.6.1 Different Models

To check the three-step estimation model presented in the last section, I conduct three alternative models and compare their regression results with those of the baseline model. In the first alternative model, I regress hours of work on mortgage status and wage variables without considering the endogeneity problem or the sample selection problems. In the second and third models, I deal with the endogeneity problem and the sample selection problem separately. The estimation results of the four models are reported in tables (3.8) and (3.9). First of all, the causal parameters are significantly positive across all models, even though the coefficients are smaller after the mortgage status is instrumented by the tax subsidy variable. Also, the coefficient for mortgage status might be much smaller before correcting the sample selection bias, which implies the importance of dealing with sample selection in this research. Second, the coefficients for hourly wage are larger after using the sample-corrected wages, which is consistent with the literature comparing first- and second-generation labor supply models. Third, other household income is significantly negative across all models, indicating a negative income effect. That is, wives are less likely to participate in the labor market or are more likely to work less when household income is high. Lastly, the number of young children and older children are both negatively associated with married wives' labor supply, which is consistent with the intuition and the literature.

3.6.2 Subsample: Does household wealth matter?

There are two mortgage related driving factors for wives to participate in the labor market and work more. The first one is active, as presented in the static labor supply model–that is, wives work more to enjoy more housing services. The second one is passive, wives are forced to work more to increase the ability to buffer the future risk, which is harmed by the mortgage commitment. Since rich households have more financial resources to support housing consumption and buffer the future risk, we expect to observe a larger effect of the home mortgage on the wife's labor supply.

To further test this hypothesis, I estimate the same three-step empirical model for two subsamples. In the first subsample, household wealth is above the median, while in the second, it is below the median. Table (3.10) presents the estimation results for the two subsamples. The causal parameter of mortgage status is larger and more significant among poor households than among rich households, which confirms the hypothesis that married women with limited household wealth are more likely to be affected by a mortgage payment.

3.7 Conclusion

This paper studies whether the female labor supply is affected by household mortgage status. The direct regression of labor supply on the mortgage variable is biased not only because I could observe wages only for those who are working, but also because the mortgage and labor supply decisions can be correlated with each other in an unobserved way. I adopt the Heckman two-step method to correct the sample selection bias. To address the endogeneity problem of mortgage status, I adopt mortgage interest deduction as an instrumental variable. The estimation shows that a home mortgage, measured by the ratio of mortgage payment to household income, has a positive impact on the female labor supply, suggesting that wives do work more to consume more housing services and to hedge potential income uncertainty in the future. Estimation for different models show that the sample selection problem might bias down this causal effect, while the endogeneity problem might bias up this effect. Lastly, the causal coefficient is larger for the subsample whose household wealth is below the median, suggesting that households with limited wealth are more likely to be affected by mortgage status.

The current research is conducted in a static setting, while housing and labor supply decisions are usually made dynamically within the household. For example, previous work experience would be positively associated with current labor supply decision. At the same time, people with longer work experience are also more likely to qualify for a mortgage loan and become a mortgagor. Therefore, it would be interesting to extend this study into a dynamic setting to investigate how these two decisions interact dynamically. In addition, this paper serves as the first step in looking at the relationship between mortgages and female labor market participation. A natural question arises: Is this effect particular to mortgage commitments, or does

Table 3.1: Summary Statistics SIPP Sample (1996-2000)

	Mean	S.D.	N
Labor participate	0.764	0.425	57318
Hours of work	121.270	68.548	57318
Mortgage Ratio	12.092	10.476	57318
Mortgagor	0.745	0.436	57318
Hourly wage	12.496	7.065	43458
Spouse hourly wage	16.694	7.772	44283
Age	42.847	10.546	57318
Number of persons in the household	3.445	1.339	57318
With children	0.564	0.496	57318
Number of young children (0-6)	0.401	0.718	57318
Number of old children (7-15)	0.742	1.007	57318
White	0.898	0.302	57318
Black	0.065	0.247	57318
High school	0.465	0.499	57318
College	0.480	0.500	57318
Graduate school	0.055	0.229	57318
Live in metro	0.750	0.433	57318
Spouse monthly income	2718.239	1410.575	57318
Total household income	4621.397	2026.034	57318
Other household income	3024.178	1548.499	57318
Total household wealth	116955.359	108706.132	57318
Observations	57318		

[1] Source: SIPP 1996 Panel
[2] Mortgage Ratio: ratio of mortgage payment to family income
[3] Other household income: household income excluding female income

it also apply to other types of debt within household. To answer this question, more analysis could be conducted to include other types of debt in the regression models.

Figure 3.1: Female labor participate rate and mortgage status across different age groups (1996-2000)

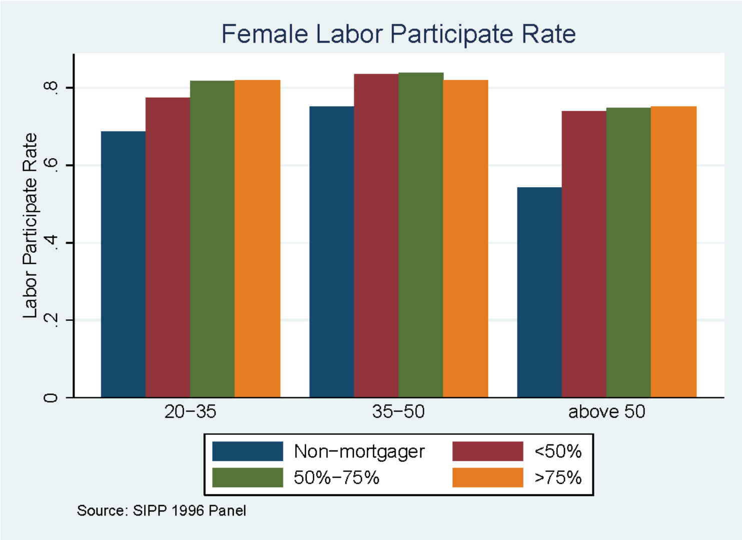


Figure 3.2: Female hours of work and mortgage status across different age group (1996-2000)

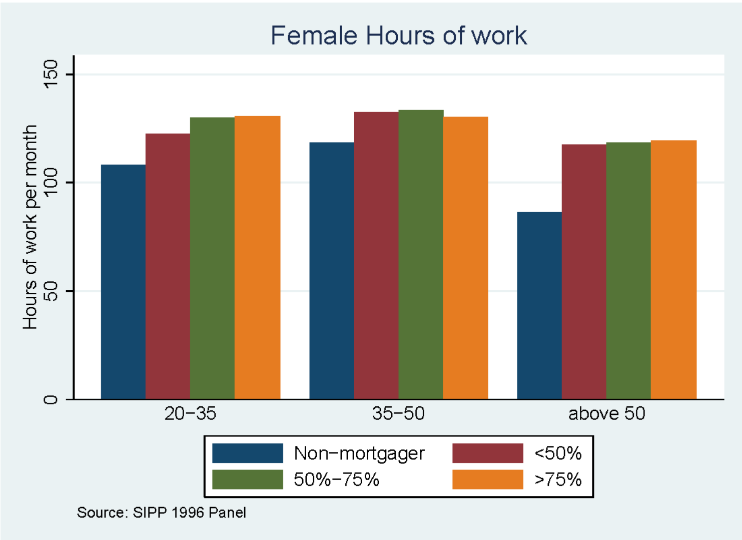


Table 3.2: NBER Mortgage Interest Subsidy Rate by US state in % (1996-2000)

Year	*State Mortgage Subsidy*	*Std. Dev.*	*Min.*	*Max.*
1996	5.21	0.14	0	10.97
1997	5.17	0.28	0	10.81
1999	4.99	0.17	0	10.4
2000	4.86	0.13	0	10.4

[1] Mortgage Interest Subsidy Rate is calculated by TAXSIM provided by NBER

Table 3.3: NBER Mortgage Interest Subsidy Rate by Year in % (1996-2000)

U.S. State	*State Mortgage Subsidy*	*Std. Dev.*	*Min.*	*Max.*
Alabama	3.95	0.31	3.65	4.28
Alaska	0.00	0.00	0.00	0.00
Arizona	3.19	0.10	3.10	3.38
Arkansas	6.69	0.00	6.69	6.69
California	6.48	0.86	6.00	8.00
Colorado	4.93	0.12	4.63	5.00
Connecticut	4.50	0.00	4.50	4.50
Delaware	6.48	0.50	5.46	6.97
DC	9.30	0.01	9.23	9.30
Florida	0.00	0.00	0.00	0.00
Georgia	5.66	0.00	5.66	5.66
Hawaii	8.98	0.68	7.81	9.37
Idaho	8.19	0.02	8.10	8.20
Illinois	3.00	0.00	3.00	3.00
Indiana	3.40	0.00	3.40	3.40
Iowa	6.81	0.41	5.89	7.47
Kansas	6.25	0.00	6.25	6.25
Kentucky	6.00	0.00	6.00	6.00
Louisiana	2.85	0.01	2.85	2.90
Maryland	4.93	0.07	4.85	5.00
Massachusetts	5.95	0.02	5.85	5.95
Michigan	4.39	0.04	4.20	4.40
Minnesota	7.78	0.35	7.05	8.00
Mississippi	4.95	0.07	4.84	5.00
Missouri	4.95	0.31	4.39	5.15
Montana	7.84	0.50	6.66	8.63
Nebraska	10.74	0.22	10.40	10.97
Nevada	0.00	0.00	0.00	0.00
New Hampshire	0.00	0.00	0.00	0.00
New Jersey	5.52	0.00	5.52	5.53
New Mexico	7.10	0.00	7.10	7.10
New York	8.38	0.11	8.17	8.44
North Carolina	7.00	0.00	7.00	7.00
Ohio	6.15	0.13	5.95	6.32
Oklahoma	6.48	0.10	6.33	6.54
Oregon	9.00	0.00	9.00	9.00
Pennsylvania	2.80	0.00	2.80	2.80
Rhode Island	5.53	1.72	3.75	7.41
South Carolina	7.00	0.00	7.00	7.00
Tennessee	0.00	0.00	0.00	0.00
Texas	0.00	0.00	0.00	0.00
Utah	6.09	0.00	6.09	6.09
Virginia	5.65	0.02	5.65	5.75
Washington	0.00	0.00	0.00	0.00
West Virginia	6.50	0.00	6.50	6.50
Wisconsin	6.84	0.10	6.55	6.93
Maine, Vermont	7.12	0.93	6.50	8.50
North Dakota, South Dakota,	0.00	0.00	0.00	0.00

[1] Mortgage Interest Subsidy Rate is calculated by TAXSIM provided by NBER

Table 3.4: First Step: Participation equation and wage equation(Heckman two-step)

	Probit	Wage equation
Age	0.160***	0.340***
	(0.005)	(0.030)
Age2	-0.002***	-0.004***
	(0.000)	(0.000)
Number of young children (0-6)	-0.423***	
	(0.010)	
Number of old children (7-15)	-0.169***	
	(0.007)	
White	0.067**	0.416**
	(0.032)	(0.169)
Black	0.362***	0.202
	(0.041)	(0.205)
College	0.287***	3.419***
	(0.013)	(0.070)
Graduate school	0.573***	8.903***
	(0.031)	(0.141)
Live in metro	0.043***	2.193***
	(0.014)	(0.073)
log(Other income)	-0.208***	
	(0.010)	
Inverse Mills ratio		-1.793***
		(0.227)
Constant	-0.097	0.764
	(0.123)	(0.686)
Observations	56786	42944
R^2		0.164
Pseudo R^2	0.114	

[1] Year dummies included, but not reported
[2] Standard errors in parentheses
[3] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.5: Second Step: Mortgage interest deduction on Mortgage Ratio

	OLS	OLS fixed effect	OLS random effect
Mortgage interest deductions(%)	0.232***	0.277***	0.144***
	(0.015)	(0.079)	(0.032)
Age	0.367***	-0.021	0.317***
	(0.032)	(0.131)	(0.058)
Age2	-0.007***	-0.002	-0.006***
	(0.000)	(0.001)	(0.001)
Number of young children (0-6)	1.383***	0.320***	0.751***
	(0.065)	(0.118)	(0.091)
Number of old children (7-15)	0.298***	0.415***	0.452***
	(0.045)	(0.113)	(0.076)
White	-2.152***		-2.393***
	(0.218)		(0.493)
Black	-3.280***		-3.825***
	(0.268)		(0.608)
College	1.113***		1.281***
	(0.085)		(0.198)
Graduate school	0.939***		1.605***
	(0.184)		(0.413)
Live in metro	4.160***	1.690***	3.902***
	(0.096)	(0.428)	(0.204)
log(Other income)	1.201***	2.580***	2.331***
	(0.049)	(0.047)	(0.044)
Constant	15.072***	36.704***	26.413***
	(0.789)	(3.686)	(1.367)
Observations	56786	56786	56786
R^2	0.135	0.061	0.134

[1] Year dummies included, but not reported
[2] Standard errors in parentheses
[3] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.6: Third Step I: Mortgage Ratio and Labor supply

	Hours of Work	Labor Participate Rate
Mortgage Ratio	0.558***	0.012***
	(0.027)	(0.001)
Predicted Hourly wage	2.891***	0.071***
	(0.106)	(0.003)
Age	7.657***	0.142***
	(0.218)	(0.005)
Age2	-0.109***	-0.002***
	(0.002)	(0.000)
Number of young children (0-6)	-19.059***	-0.410***
	(0.435)	(0.010)
Number of old children (7-15)	-7.907***	-0.167***
	(0.296)	(0.007)
White	1.254	0.026
	(1.446)	(0.032)
Black	12.545***	0.321***
	(1.775)	(0.041)
log(Other income)	-7.055***	-0.227***
	(0.322)	(0.010)
Constant	42.265***	0.027
	(5.223)	(0.124)
Observations	56786	56786
R^2	0.128	
Pseudo R^2		0.119

[1] Year dummies included, but not reported

[2] Mortgage Ratio: ratio of mortgage payment to family income

[3] Other income is defined as household income net of wife's and spouse's income

[4] Standard errors in parentheses

[5] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.7: Third Step II: Mortgage Ratio and Labor supply (Spouse's income)

	Hours of Work	Labor Participate Rate
Mortgage Ratio	0.310***	0.007***
	(0.059)	(0.001)
Predicted Hourly wage	3.658***	0.088***
	(0.225)	(0.006)
Age	7.606***	0.132***
	(0.551)	(0.012)
Age2	-0.105***	-0.002***
	(0.006)	(0.000)
Number of young children (0-6)	-15.168***	-0.313***
	(1.048)	(0.022)
Number of old children (7-15)	-6.501***	-0.131***
	(0.547)	(0.012)
White	3.098	0.066
	(2.350)	(0.052)
Black	7.049**	0.189***
	(3.036)	(0.069)
log(Spouse income)	-3.119***	-0.081***
	(0.586)	(0.015)
log(Other income)	-0.877***	-0.022***
	(0.277)	(0.007)
Constant	-4.222	-1.260***
	(13.352)	(0.293)
Observations	14270	14270
R^2	0.102	
Pseudo R^2		0.093

[1] Year dummies included, but not reported

[2] Mortgage Ratio: ratio of mortgage payment to family income

[3] Other income is defined as household income net of wife's and spouse's income

[4] Standard errors in parentheses

[5] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.8: Robustness check I: Different Models

Hours of work

	Baseline	Check 1	Check 2	Check 3
Mortgage Ratio	0.558***	0.218***	1.202***	0.066***
	(0.027)	(0.083)	(0.273)	(0.009)
Predicted Hourly wage	2.891***	0.113***	2.571***	0.111***
	(0.106)	(0.011)	(0.107)	(0.011)
Age	7.657***	0.671***	7.506***	0.682***
	(0.218)	(0.068)	(0.219)	(0.067)
Age2	-0.109***	-0.008***	-0.106***	-0.008***
	(0.002)	(0.001)	(0.002)	(0.001)
Number of young children (0-6)	-19.059***	-0.859***	-19.939***	-0.804***
	(0.435)	(0.132)	(0.435)	(0.132)
Number of old children (7-15)	-7.907***	-0.413***	-8.087***	-0.408***
	(0.296)	(0.088)	(0.296)	(0.088)
White	1.254	-0.202	3.110**	-0.374
	(1.446)	(0.422)	(1.449)	(0.423)
Black	12.545***	0.063	14.930***	-0.135
	(1.775)	(0.508)	(1.780)	(0.508)
log(Other income)	-7.055***	-0.153*	-6.514***	-0.232***
	(0.322)	(0.087)	(0.322)	(0.087)
Constant	42.265***	146.246***	31.556***	147.378***
	(5.223)	(1.565)	(5.221)	(1.572)
Observations	56786	42511	56786	42511
R^2	0.128	0.007	0.122	0.008

[1] Check 1 doesn't deal with problems of sample selection and endogeneity of mortgage status

[2] Check 2 deals with only the problem of sample selection

[3] Check 3 deals with only the problem of endogenity of mortgage status

[4] Year dummies included, but not reported

[5] Standard errors in parentheses

[6] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.9: Robustness check I: Different Models
Labor participation

	Baseline		Check 1		Check 2		Check 3	
	β / SE	Mfx	β / SE	Mfx	β / SE	Mfx	β / SE	M
Mortgage Ratio	0.012***	0.003***	0.062***	0.001***	0.027***	0.008***	0.014***	0.00
	(0.001)		(0.020)		(0.006)		(0.002)	
Hourly wage	0.071***	0.021***	0.053***	0.001***	0.063***	0.018***	0.052***	0.00
	(0.003)		(0.004)		(0.003)		(0.004)	
Age	0.142***	0.041***	0.118***	0.002***	0.138***	0.040***	0.122***	0.00
	(0.005)		(0.014)		(0.005)		(0.014)	
Age2	-0.002***	-0.001***	-0.001***	-0.000***	-0.002***	-0.001***	-0.001***	-0.00
	(0.000)		(0.000)		(0.000)		(0.000)	
Number of young children (0-6)	-0.410***	-0.119***	-0.161***	-0.003***	-0.426***	-0.124***	-0.151***	-0.00
	(0.010)		(0.028)		(0.010)		(0.028)	
Number of old children (7-15)	-0.167***	-0.048***	-0.077***	-0.001***	-0.171***	-0.050***	-0.075***	-0.00
	(0.007)		(0.020)		(0.007)		(0.021)	
White	0.026	0.008	-0.052	-0.001	0.060*	0.017*	-0.085	-0.0
	(0.032)		(0.107)		(0.032)		(0.108)	
Black	0.321***	0.093***	-0.012	-0.000	0.362***	0.105***	-0.047	-0.0
	(0.041)		(0.130)		(0.041)		(0.131)	
log(Other income)	-0.227***	-0.066***	-0.070**	-0.001**	-0.214***	-0.062***	-0.098***	-0.00
	(0.010)		(0.029)		(0.010)		(0.030)	
Constant	0.027		0.392		-0.186		0.655*	*
	(0.124)		(0.348)		(0.123)		(0.355)	
Observations	56786		42511		56786		42511	
Pseudo (R^2)	0.119		0.075		0.113		0.084	
LR chi2	7402.912		356.372		7037.124		402.108	
Prob > chi2	0.000		0.000		0.000		0.000	

[1] Check 1 doesn't deal with problems of sample selection and endogeneity of mortgage status
[2] Check 2 deals with only the problem of sample selection
[3] Check 3 deals with only the problem of endogenity of mortgage status
[4] Year dummies included, but not reported
[5] Standard errors in parentheses
[6] * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.10: Robustness check II: Does Household Wealth Matter?

	Baseline	Household wealth>median	Household wealth<median
Mortgage Ratio	0.558***	0.269***	0.807***
	(0.027)	(0.036)	(0.043)
Predicted Hourly wage	2.891***	2.858***	3.007***
	(0.106)	(0.145)	(0.161)
Age	7.657***	5.445***	11.760***
	(0.218)	(0.286)	(0.417)
Age2	-0.109***	-0.083***	-0.151***
	(0.002)	(0.003)	(0.004)
Number of young children (0-6)	-19.059***	-18.303***	-18.915***
	(0.435)	(0.511)	(0.814)
Number of old children (7-15)	-7.907***	-7.601***	-7.385***
	(0.296)	(0.363)	(0.510)
White	1.254	1.048	0.791
	(1.446)	(1.864)	(2.274)
Black	12.545***	10.484***	13.354***
	(1.775)	(2.178)	(3.274)
log(Other income)	-7.055***	-6.244***	-8.980***
	(0.322)	(0.421)	(0.501)
Constant	42.265***	79.204***	-46.202***
	(5.223)	(6.591)	(10.333)
Observations	56786	32990	23796
R^2	0.128	0.114	0.150

[1] Year dummies included, but not reported
[2] Standard errors in parentheses
[3] $^{*}p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 4

The Impact of Labor Migration on Children's Health: Evidence from Rural China

4.1 Introduction

The changing economic climate in China has caused a dramatic increase in 'labor migration'. Labor migration is the migration of Chinese rural residents to bigger cities where higher-paying, temporary jobs are available. In 2009, the floating population in China reached 211 million adults, leaving over 58 million children behind in homes far from their parents.

The utility function of parents is composed by household consumption, children's health and education. The main reason for labor migration is to improve household financial situation. As a result, they could increase household consumption, afford better education for their children, and better health insurance. Conventional wisdom suggests that these left-behind children are at risk of developing health problems and physical and psycho-social stress[1]as a result of a lack of parental guidance and relevant health information. These issues raise concerns for social workers and policy makers. Nevertheless, despite the fact that migrated parents are spending less time with their children, these parents are able to provide better remittances,

[1] Currently, the schools in rural China do not have the adequate systems or a relevant curriculum in place to address these issues.

nutrition and health relevant information as a result of their increased income and the knowledge they obtain through their migration experiences. Little is known about the extent to which the health of left-behind children is affected in China, particularly those children who are too young to take care of themselves.

This paper aims to establish the overall consequences of parental migration on the health outcomes and childcare of their left-behind children. The data used in the analysis are primarily derived from four waves of the China Health and Nutrition Survey (CHNS), collected in 2000, 2004, 2006 and 2009. The CHNS was designed to examine the effects of Chinese health, nutrition, and family planning policies. The people of nine provinces that vary substantially in geography, economic development, and access to public resources were surveyed.

Some of the economic literature that focuses on labor migration in China suggests that the remittances forwarded to families by migrated members benefit the households financially. For instance, Du, Mroz, Zhai, and Popkin (2004) and De Brauw (2008) found that labor migration increases family consumption level. Giles (2006) also found that having migrated family members could improve the family's risk-coping ability. On the other hand, there are also papers that focus on the left-behind family members, particularly school-age children. Chen, Huang, Rozelle, Shi, and Zhang (2009) found that educational outcomes of children improved in migrant households. However, De Brauw and Mu (2011) found that the nutrition of some school-age children from migrant households was negatively affected.

There are a few papers that study the health outcomes of left-behind children in China. One of them is Mu and van de Walle (2011), which examined the weight of left-behind children, and found that older children (7-12 years) were more likely to be underweight in migrant households than those who lived in non-migrant household. Zhang (2012) used survey data from the 2000 wave of the CHNS to study the impact of labor migration on children's health. She found no significant health outcome effects for children whose fathers had migrated. Both papers, however, do not consider the potential endogeneity of parents' migration and children's health. Therefore their results might be biased.

There is growing evidence that parents' socio-economic status would affect children's intellectual, health and behavioral development at early ages and these early development casts a long shadow over subsequent achievements (Cunha and Heckman (2007), Heckman (2008), and Heckman, Moon, Pinto, Savelyev, and Yavitz (2010)). Most of the evidences found so far are from developed countries. This paper contributes to the literature by exploring the early child development in China where the economy is going through a dramatic change and the early childhood interventions is still quite limited.

The main methodological obstacle of quantifying the effect of parent's migration is the endogeneity problem. This may be manifested as a problem of reverse causation. Instead of being affected by parents' migration status, children's health status could be a critical factor for parents when making migration decisions. For example, parents whose children are in poor health may have to stay home to take care of their children. On the other hand, they may have stronger financial incentive to migrate to earn extra money to finance better medical care for their sick children. Moreover, parents' migration decisions could be correlated to children's health through unobserved variables, such as genetically inherited health deficiency, whereby sick parents would be too sick to leave their sick children and migrate to urban areas for work. Therefore, the significant correlation between parents' migration and children's health status may not indicate causality.

To solve the endogeneity problem, we use instrumental variables (IV) estimation. To be more specific, we instrumented father's migration status with the average male migration rate, using historical county data, instrument mother's migration status with historical county level female migration rate, and instrument household migration status with historical county level household migration rate. The historical county level migration rate is calculated as the average local migration rate from the previous survey year. The historical county level migration rate by gender is a suitable indicator to reflect the local culture and network of migration, where the network refers to a person's exposure to migration information from her migrated friends or family members. Intuitively, people living in the areas with a tradition of migration or with a better migration network are more likely to migrate. In the first stage regression of this paper, it can be seen that this set of instruments have strong predictive power on parents' migration status. One might be concerned that these instruments could influence child health directly, since county level migration rates are also correlated with the local average income level. To address this, I included the county level average income as an explanatory variable.

In this paper, we adopted the panel structure of the data and employed a fixed effects model to study the overall health status of left-behind children. The causality effect of migration is identified by two-stage estimation. The estimation results are presented with and without the IV correction. Moreover, we conducted two robustness checks to support our estimation results. In the first robustness check, we excluded household income as an explanatory variable, as household income might be correlated with unobserved shocks that could also affect children's health. This correlation could lead to biases in estimation. In the second robustness check, I excluded the number of elders in the household, as family size could affect peoples' migration decisions because children could be taken care of by other family members. As a result, the estimation results might be biased.

Generally speaking, we found there were few significant effects of parents' migration on child outcomes. A possible explanation for this is that the coefficients capture the net effects of parents' migration. Children with migrated parents receive less physical care, but may receive more financial support, access to better nutrition products sent from their mothers, and better nutritional information. There are both positive and negative effects on children's health. The coefficients imply that the positive effects of parents' migration are about the same as the negative effects on children's health. Though the regression results on the whole sample were not significant, the regression results from subsamples provided more insights. It showed that children aged between 5 and 10 are positively affected by fathers' migration, possibly because these children received higher remittance, better access to nutrition information and products.

Our paper contributes to the literature in a number of ways. Firstly, we used novel instrumental variables dealing with the endogenous nature of parents' migration decisions, which are able to predict the migration propensity of parents. Secondly, we studies different causality effects of father's and mother's migration status on children's health outcomes, which were significantly different. Thirdly, in addition to traditional measurements of child health that focus on height and weight, we also considered nutrient intake (consumption of calories and protein), immunization shots and childcare. These measures provided a more comprehensive picture of the impact of labor migration on children's health.

The paper proceeds as follows: Section 2 discusses the history of labor migration and child nutrition in China; Section 3 describes the conceptual framework; Section 4 discusses the data; Section 5 describes the empirical specification; Section 6 presents the main results regarding the effect of parent migration on the physical health of children; Section 7 goes through several robustness checks; Section 8 discusses the results from subsamples; and a conclusion is provided in Section 9.

4.2 Background

According to the analysis report of labor migration in China by the National Bureau of Statistics of China (2012)[2], the total number of migrated labor from rural areas increased from 225 million in 2008 to 252 million in 2011. The rapid growth of rural-to-urban migration has been an important demographic trend in China. In this section, we first introduce the background of labor migration in China and its impact on rural communities, followed by a discussion on how the heath of rural children has changed over time.

4.2.1 Labor Migration and Children Left Behind in Rural China

Since 1958, under the central planned economy in China, China has used the household registration system (HuKou system) to control the labor migration from rural to urban areas. Under the HuKou system, households are divided to Agriculture HuKou and non-Agriculture HuKou, where the rural-urban migration was strictly restricted. In the 1990s, 83% of households were classified under the Agriculture HuKou category, according to Mallee (1995).

In 1988, the HuKou reform took place, whereby rural migrants were allowed to obtain a temporary residence. However according to a World Bank report[3], rural migrants were not able to access the urban welfare system, including education, health and the social safety net. Therefore the rural migrants maintained a close tie to their hometown village, as their benefits were linked to their household registration status.

According to Bao, Bodvarsson, Hou, and Zhao (2009), the large income gap between urban and rural areas, created by decades of urban-rural segregation and uneven economic growth, provided strong incentives for rural people to move to

[2]See National Bureau of Statistics of China. 2012. "Year 2011 Report on the Rural-Urban Labor Migration in China." stats.gov.cn, $http://www.stats.gov.cn/tjfx/fxbg/t$20120427_402801903.$htm$. for more details.

[3]From Poor Areas to Poor people: China's Evolving Poverty Reduction Agenda (2009).

urban areas, especially after rural-urban labor flow was officially permitted. As a result, China has experienced dramatic changes in its labor market since the 1990s. Liang and Ma (2004) found that the migration population grew from 20 million in 1990 to 45 million in 1995 and to 79 million in 2000 using the one percent sample from the 1990 and 2000 waves of the Population Census and one percent sample from the 1995 wave of population survey.

It is important to note the different migration rates by gender, as mother's migration may have different impact on child health than father's migration. According to Zhao (1999) and Rozelle, Guo, Shen, Hughart, and Giles (1999), there were substantially more migrated men than women in the mid-1990s. Mu and van de Walle (2011) showed that the gender gap in migration has increased over time. Our findings using CHNS data support this.

4.2.2 Health of Children in China

The health of children in China has improved with economy growth. Shen, Habicht, and Chang (1996) showed that the average height of children aged two to five years had increased by 3.8 cm in 1990 when compared with data from 1975. Chen (2008) found the prevalence of underweight children and the rate of stunting (the percentage of children with Height-for-age Z-scores below two) among Chinese children declined from 1990 to 1995. Svedberg (2006) found that the stunting rate had decreased further by 2002. Additionally, Osberg, Shao, and Xu (2009) showed that height-for-age Z-score in children increased between 1991 and 2000. The changes in children's health might be explained by the improvement of the diet quality in China, which is supported by Du, Mroz, Zhai, and Popkin (2004). They showed that the nutritional intake of children shifted from carbohydrates to high fat and high energy-density foods.

Although the health of children in China has improved on average, malnutrition is still an issue. According to De Brauw and Mu (2011), the stunting rate in 2002 was still nearly 15%, indicating a substantial portion of the population remain malnourished. There are also other challenges in improving nutrition among children. Liu, Fang, and Zhao (2013) analyzed urban-rural disparities of China's child health and nutritional status using CHNS data from 1989 to 2006, and showed that on average, urban children have 0.29 higher height-for-age z-scores and 0.19 greater weight-for-age z-scores than rural children.

4.3 Conceptual Framework

There are at least three main channels through which migration might affect the health status of children: the income effect, the time effect, and the information effect.

First of all, the primary reason for a member of a household to migrate is to increase household income. We anticipate the increased family income will have a positive effect on child health outcomes for various reasons. For example, extra income could increase diet quality (Du, Mroz, Zhai, and Popkin (2004)), by switching from high carbohydrate food to high fat and high energy-density foods. Therefore, the calorie intake may increase when income increases. Moreover diet improvements might improve height-for-age Z-score and weight-for-age Z-score. Finally, health service utilization for children may increase as well. For example, migrant parents may be able to afford to have their children immunized as a result of increased income.

The second channel through which migration may affect the health status of children is through the time allocated to child care. Mu and van de Walle (2011) found that when one family member leaves for urban work, the remaining family members must take on an increased farm work load. As a result, they may spend less time cooking and child rearing. Consequently, child health outcomes may be affected. In cases where both parents have migrated, children might be left in the care of relatives, usually their elderly grandparents. In such cases, children might not have a regular diet routine and may eat poorly. As a result, the child's nutrient intake, and subsequently, their height and weight, may be affected.

The third channel is though better access to nutritional information from migrated parents. People always migrate to urban areas that have better economic conditions and health services. Therefore migrants should have better access to nutritional information. For example, migrants may learn more about healthy diets, and encourage their children to eat more nutritious foods. Moreover, they may learn more about the importance of immunization, and have the incentive to let their children get immunized.

As explained above, the direction of the effect of parent migration on child health outcomes is ambiguous. In the next section we present the data and empirical framework.

4.4 Data

The China Health and Nutrition Survey (CHNS) was designed to examine the effects of the health, nutrition, and family planning policies and programs implemented by national and local governments and to check how the social and economic transformation of Chinese society is affecting the health and nutritional status of the Chinese population. The Survey covered nine provinces that vary substantially in geography, economic development, and access to public resources. Demographic characteristics, household assets and other information were also collected as part of the survey. The first round of the CHNS, including household, community, and health/family planning facility data, was collected in 1989. Seven additional panels were collected in 1991, 1993, 1997, 2000, 2004, 2006 and 2009.

From 1997 onwards, families were asked to provide reasons for migrated or absent family members as part of CHNS. A migrant was defined as any individual who had left the home at the time of the survey to seek employment. The data used in the analysis were primarily derived from four waves of the CHNS, collected in 2000, 2004, 2006 and 2009. The reason that we did not use data from the 1997 wave of the survey is because we used the historical migration rate from the previous wave as instrument variables, and this information was not available for the 1997 wave.

In the first wave (1997) of the CHNS, 15,917 individuals were surveyed. Survey response rates and attrition are difficult to determine for two reasons: firstly, the participants who had migrated in one survey year may have returned home in a later year; and secondly, new participants were recruited following the 1997 survey, to replenish samples if a community had less than 20 households, or if participants had formed a new household or separated from their family into a new housing unit in the same community. If we calculated response rate based on those who participated in previous survey rounds remaining in the current survey, our response rates would be around 88% at individual level and 90% at household level (Popkin, Du, Zhai, and Zhang (2010)). Mu and van de Walle (2011) showed that the attrition was random and should not generate panel attrition bias.

The CHNS provides rich information on parents' migration status and children's health outcomes, which make it possible to analyze the correlation between the two. However, this data provides limited information for the purpose of testing the mechanisms behind the effect of parents' migration. For example, we can't observe the how long the parents have been away from home, how much money they send to home or how much time is spend on child care every month, which makes it impossible to distinguish and quantify the size of income effect and time effect.

For estimation purposes, we dropped observations where one or more of the following critical pieces of information pieces were missing: (a) child's height, (b) child's weight, (c) child's calorie intake, (d) child's protein intake, (e) parents' education level, and (f) parents' migration status. To calculate the height-for-age Z-score (HAZ) and weight-for-age Z-score (WAZ), we used the most recent growth charts made available by the World Health Organization (WHO). To measure child's calorie and protein intake, we used a set of age and gender-specific Recommended Dietary Allowances (RDAs) sanctioned by the Chinese Nutrition Society (2000). R-DAs are based on average energy allowances, i.e. calorie intake for each specific age and gender group.

We randomly selected one child from families with several children to avoid any biases of related children and other unobserved variables. In this paper, we focus on children under ten years of age, because they are at greater risk of developing problems associated with malnutrition and are more likely to respond to nutritional interventions (WHO, 1995). We excluded households in which the children were older than ten. After the aforementioned data altering procedures, the final sample data is unbalanced panel data, containing 1,600 children and 2,201 observations.

There are several reasons that only 40% of the children had more than one observation in the data. The first is that we only kept the observation when we had both the child data and their parent's data. For instance, if the mother or father did not respond to the survey, the child's response was excluded as it could not be used. As the individual response rate is 88%, the probability that the child is included in the next survey year is calculated by multiplying the child's response rate by their parents' response rates, which equates to 0.68 (0.88^3). The second is the individual response rate is not 88% for each survey year - it is 83% in year 2000, and 80% in year 2004 (Popkin, Du, Zhai, and Zhang (2010)). The third is that there are missing variables. For instance, the response rate of the question for migration status is less than 80%. After the exclusion of children who are younger than 10, there is approximately 40% probability that a child is included in more than one survey wave of the survey.

From table (4.1), we can see the migration rate kept increasing and reached a peak at year 2006. The table shows that fathers were more likely than mothers to migrate from households. Both parents had migrated from relatively few families, implying that most families had one parent left in the household to take care of the children. From the data, it is clear that labor migration became quite common in rural areas after year 2000. In year 2006, 21% of children had at least one parent who had migrated, and both parents of 4% of the children sampled had migrated. However these figures likely underreport the true scale of migration because we did not account for migration that took place over shorter periods of time (Cai, Park, and Zhao (2004)).

Since the migrate rate in our sample is relatively small. To make sure our sample size is large enough to measure migration rate, we conduct power analysis for migration. The default significance level (alpha level) is set at 0.05 and the power is set to be 0.9. Taking father's migration, which is 15% in the sample, as an example. We find the sample size that need to test whether father's migration rate is different from 13% or 17% are 354 and 378, which are smaller than our sample size. The power analysis implies our sample size is large enough to measure parent's migration rate.

Table (4.2a) compares differences in health outcomes and care of children between children with and without migrated parents. Children are defined as left-behind if one of their parents was a migrant. According to the table, the left-behind children on average consumed less protein than children who lived with both of their parents. At the same time, left-behind children were shorter and weighed less on average than children who lived with both of their parents. Table (4.2b) shows differences between children with and without migrated fathers in health outcomes and care. By comparing the data from Table (4.2a) and (4.2b) it is evident that there were fewer significant differences of child health outcomes and care for families with migrant fathers and non-migrant fathers. Children with migrant fathers have significantly smaller weight-for-age Z-score and protein/RDA. Table (4.2c) shows the differences between children with and without migrated mothers. Unlike children with migrant fathers, children with migrant mothers consumed significantly less protein and calories. Although the rates of migration were smaller for mothers, they seemed to have more of a significant effect on child outcomes than father migration or household migration.

We can also see that for both migrant and non-migrant households, the average height-for-age Z-score and weight-for-age Z-score were less than 0. The z scores show that children in China are on average shorter and lighter in weight compared to the WHO standards. The WHO standards were formulated in the 1970s by combining growth data from two distinct data sets in USA. The summary statistics show that children in China have relatively poor health conditions compared to the children in USA, while left-behind Chinese children are even more disadvantaged compared to Chinese children who live with both parents. Moreover the average Calories/RDA and Protein/RDA ratios are under 1 for both migrant and non-migrant households, which implies that children in China on average consume less protein and calories than recommended.

Table (4.3) shows the summary statistics of the control variables. Household income is lower in households with migrants. The difference in income could be explained by the fact that the migrated household members' income is not included in household income, although the remittances provided by the migrant are included. The table also shows that migrated parents have lower education level and are younger. This trend could be a result of the local economic conditions, as people who live in areas with better economic conditions are less likely to migrate. They also tend to receive more education and have children later in life. For similar reasons, county level average height and weight are lower for migrant households because they are proxies for features of local economy development. Moreover the number of females over 60 in the household is higher in households with migrants, which suggests that the number of elders in the household influences families' migration decisions. In general, people who migrate are more likely to live in big families, and poor areas. At the same time, they are more likely to have lower education levels and have children at younger ages. The historical county level migration rates will be used as instrumental variables and will be discussed later.

4.5 Empirical Specification

In this paper, we adopt three sets of measures of health status. The first includes child's weight-for-age Z-score (WAZ), height-for-age Z-score (HAZ). The second set includes child's daily calorie intake, child's daily calorie intake/RDA, child's daily protein intake, and child's daily protein intake/RDA. The third set includes the number of immunization shots that the child received in the survey year, and whether the child has been cared for by non-household members.

We aimed to identify cause-effect relationships of parents' migration status on children's health outcomes. In addition to parents' migration status, child health is also affected by other demographic factors, such as gender, parents' education level, family size, the number of siblings, and household income. These were used as control variables in the estimation model.

With panel data, two models could be applied: the fixed effects model or random effects model. The Hausman test showed that the random effects model is inconsistent. The fixed effects model is employed in this paper. The panel data is unbalanced. There are 480 children with more than one observation in this data set, which is the effective sample. Among the effective sample, there are 112 parents who changed their migration status. The number of parents who changed their migration status in different survey years helped us to identify the impact of migration on children's health.

We employed three separate fixed effects models to identify the effects of household migration, fathers' migration and mothers' migration on child health outcomes and care. The fixed effects model that we employed to identify the effect of household migration

$$H_{it} = \alpha_i + \beta_1 M_{it} + \beta_2 X_{it} + \epsilon_{it} \tag{4.1}$$

where H_{it} is child i's health outcome at time t, M_{it} is child i's household migration status at time t. The dummy variable equals to 1 if either or both the child's parents had migrated out at time t, and 0 otherwise. X_{it} is a vector of demographic variables including gender dummy (female as 1), parents education level, household income, the number of males aged over 60 in the household, the number of females aged over 60 in the household, the number of boys under age ten in the household, the number of girls under age ten in the household, the county level average height, the county level average weight, the county level average daily calorie consumption/RDA, the county level average protein consumption/RDA, and the county level average income. Here ϵ_{it} is an error terms for individual i at time t. [4]

The fixed effects models that we used to identify the effect of father's and mother's migration on child health are similar to Equation (4.1). The only difference is the dummy variable M_{it}. To capture the effect of fathers' migration, the dummy

---

[4] To control for the trend effect of migration, year dummies can be added to the regression model for a trend analysis.

variable M_{it} is redefined to equal to 1 if the child's father has migrated out at time t, and 0 otherwise. To capture the effect of mothers' migration, the dummy variable M_{it} is redefined to equal to 1 if child's mother has migrated out at time t, and 0 otherwise.

We did not include the number of working age males/females in the household as explanatory variables for two reasons. The first is that we have already controlled the household income and parents' migration status. The second is the preliminary results show that the number of working age males/females in the household does not have a significant effect on children's health outcomes. In the model, we use the number of boys/girls instead of the number of siblings because many children come from large families in rural China and often live with their cousins and their siblings. Therefore the total number of children in the household could impact the child's health.

Household income is used as a control variable instead of individual income. The reason is that there are too many missing values for individual income, especially for migrants. The remittances are included in household income but we cannot break them out, as the survey did not ask about the amount of remittances. We included more variables that measure the households' assets as explanatory variables, but the coefficients are not significant. Finally, we only kept household income in the regression.

Plenty of literature mentioned the biases that may be caused due to the endogenous nature of labor migration. In our CHNS sample, endogeneity mainly arose because a child's health status also affects parents' migration decisions. The common methodology adopted to correct such biases has been used as an instrumental variable approach, isolating exogenous variation in parents' migration status. We adopted an IV approach and used historical county level migration rates as instruments. The historical county level migration rate is calculated as the local migration rate from previous survey year. The historical migration rate could proxy the migration network. The difference between the average male migration rate and female migration rate could also be a proxy for local culture.

4.6 Estimation Results

4.6.1 Results of Ordinary Least Squares model

As a baseline, Table (4.4a) and Table (4.4b) present the baseline effects of the household migration status on child health outcomes and care from the ordinary least squares regressions. Here, the child household migration status dummy variable equals one if either or both the child's parents have migrated. Table (4.5a) and Table (4.5b) present the effects of the fathers' migration status on child health outcomes and care from the ordinary least squares (OLS) regressions. Table (4.6a) and Table (4.6b) present the effects of the mothers' migration status on child health outcomes and care from the ordinary least squares regressions.

Though the OLS regression analysis may not be able to capture the exact relationship between labor migration and children's health, the results give us an idea of the correlation between children's health and the explanatory variables. It shows that parental migration does not necessarily negatively correlate with children's health outcomes. Firstly, coefficients are similar for father's migration and household migration status because the majority of household migrations are fathers' migration. Father's migration and household migration are positively correlated with children's height-for-age Z-score, and negatively correlation with the number of immunization shots that children received. However father's migration and household migration have no significant correlation with children's nutrient intake. Secondly, compared with father's migration, a mother's migration has a higher correlation with children's health outcomes, although the rate of migration is smaller for mothers. For instance, mothers' migrations is positively correlated with children's height-for-age Z-score and negatively correlated with children's daily calorie and protein intakes. The fact that migrated mothers are more likely to access child care knowledge may explain this correlation, as childcare knowledge is positively correlated with children's physical outcomes. However, a mother's absence from home means they are not able to pay attention or take care of their child's diet, which leads to lower calorie and protein consumption in their children.

When the OLS regression results are compared to Table (4.2a), Table (4.2b) and Table (4.2c),the coefficients of parents' migration and household migration cease to be significant for some measures of child health outcomes and care in the OLS results. This may be because both parents/household migration and children's health outcome are correlated with the added explanatory variables in the OLS regression. For instance, children's weight-for-age Z-score is significantly different for migrant household and non-migrant household in Table (4.2a), but the coefficient of household migration on children's weight-for-age Z-score is not significant in Table (4.4a). It can be seen that in Table (4.4a) children's weight-for-age Z-score is significantly correlated with fathers' education, county level average weight and height. At the same time, we can see from Table (4.3) that fathers' education, county level average weight and height are all significantly different for migrant household and non-migrant household. Therefore the correlation between those control variables and household migration status explains the difference in the OLS results and the summary statistics. Unlike father's migration and household migration, mother's migration remain significant in Table (4.6a) and Table (4.6b) for the variables that are significantly different for children with migrant mothers and non-migrant mothers in table (4.2c). The correlations between mothers' migration status and some child outcomes remain significant when variables are added.

4.6.2 Results of Fixed Effects model

Table (4.7a) and Table (4.7b) shows the effects of household migration status on the health outcomes and care using the fixed effects model approach without considering the endogeneity of migration. Similarly, Table (4.8a) and Table (4.8b) show the effects of father's migration status on children's health outcomes and care. Table (4.9a) and Table (4.9b) show the effects of mother's migration status on children's health outcomes and care.

With the aid of the fixed effects model, we considerably reduced the threat of omitted variable bias. From the OLS regression results we can see that most of the coefficients of parents' migration and household migration become insignificant in the fixed effects model results, especially the coefficients of mothers' migration. The results imply that there must be some omitted variables that are correlated with parents' migration decisions and may have casual effects on children's outcomes. Even though we have tried to include most of the relevant variables for children's outcome, due to the limitations of the data available, some factors may still be left out. For instance, we cannot observe whether the child has a chronic health condition. Chronic health conditions are defined as a health problems that persist for over three months, affects the child's normal activities, and require hospitalization and/or home health care and/or extensive medical care [5]. Children with chronic health conditions usually require more time and care from their parents, as well as increased financial support. Within Chinese families, the mother usually spends more time taking care of the child while father is the main income provider. Therefore, in households with a chronically ill child, compared to households with healthy children, the mother is more likely to stay at home (less likely to migrate), while father is more likely to migrate for

[5] such as Asthma (the most common) and Sickle cell anemia

higher wages. For the above reason, the results from the fixed effects model show that household migration and father's migration are now negatively correlated with children's weight-for-age Z-score, and they are not significant in the OLS model. For the same reason, the coefficients of mothers' migration become insignificant in the fixed effects model results.

4.6.3 Results of Fixed Effects model with instrument variable

Besides omitted variable bias caused by children with chronic health conditions, endogeneity bias may be partially responsible for the insignificant fixed effects results. First of all, the endogeneity could be a result of reverse causality. Parents' migration decisions may depend on children's health status. For instance, mothers are less likely to migrate when children have relatively poor health status. Moreover both parents' migration decisions and children's outcomes could be correlated with local environment and development level. Though we have tried to control those local factors by adding county average variables such as income as independent variables, it is hard to control all the local differences using current data. For example, the available data provided little information on the availability and condition of local transport. In towns that have railways or paved roads, people are more likely to migrate, and the local market is more prosperous, factors which could favor children's health. In this case, both parent migration and children's outcome are positively correlated with these unobservable factors, which may strengthen the positive correlation between them.

To solve this endogeneity problem and identify the potential causality effects of migration on children, we adopted the instrument variable method. The three endogenous variables are the household migration status dummy variable, the father's migration status dummy variable and the mother's migration status dummy variable. The child's household migration status dummy variable equals to one if either or both their parents have migrated. The instrumental variables are the historical county level average household migration rate, the historical county level average male migration rate, and the historical county level average female migration rate respectively. The instruments are gender specific. In the survey data, there are between 20 to 30 households in each county. The instruments capture the migration network and local culture. It is conceivable that the migrant network affects migration decisions. From Table (4.3), we can see that households with higher historical migration rates are more likely to have migrant household members. The local average migration rates may affect children's health and care as a result of the income the parent earned from the urban job. Once we control for the household income directly in the regression, the local average migration rate is unlikely to affect children's anthropomorphic outcomes. Another threat to the validity of the IV is that both IV and children's health outcome may be correlated with unobserved variables. For instance, the government policy may affect both the historical migration rate and children's health. In China, the change of the HuKou system is the biggest change in government policy that affects labor migration. The policy may affect children's health through the development of the local economy and labor migration. As we have already controlled the county level average income in the regression, the change of the HuKou system is unlikely to effect children's health.

Table (4.10) presents the first-stage results from the fixed effects regression. The historical average migration rate is strongly correlated with individual and household migration status. We have calculated the F-statistics against the null that the excluded instruments are irrelevant. The F-statistics are 6.65, 5.09 and 7.19 on 1 and 821 degrees of freedom for historical county level male migration rate, historical county level female migration rate, and historical county level household migration rate respectively. A common rule of thumb for models with one endogenous regressor is: the F-statistic against the null that the excluded instruments are irrelevant in the first-stage regression should be larger than 10 (Stock, Wright, and Yogo (2002)). The instruments we use are not strong instruments, and it may cause bias towards the OLS estimator. However the coefficients estimated on the instrumental variables are still significant at the five percent level, which shows the predictive power of the historical county level migration rate. As the average migration rate is a measure of the local migration network, the regression results support the hypothesis that the local migration network is a crucial factor that affects individuals' migration decisions in the corresponding local area.

The effects of household migration status on children's health outcome and care from the fixed effects model using the IV approach are presented in Table (4.11a) and Table (4.11b). The effects of father's migration status on children's health outcome and care from the fixed effects model using the IV approach are presented in Table (4.12a) and Table (4.12b). The effects of mother's migration status on children's health outcome and care are presented in Table (4.13a) and Table (4.13b).

After the correction of the endogeneity, there are few significant effects of parents' migration on children's outcomes. There are three possible reasons. The first is that the IV approach removes the reverse causality between parents' migration and children's health. The second is that the weak instrument we use may cause bias toward the OLS estimator. As a result, household migration and father' migration may lead to an even higher increase in children's weight than reported in the tables. The third is that IV usually reduces significance. It is not surprising that after applying the IV, more coefficients became insignificant. So correcting for endogeneity did not change the results.

We have discussed that parental migration effects a child's health in three major ways: the income effect, the allocation of time and the information effect. As we have controlled the income effect by including household income as one of the explanatory variables, the net effect of the parents' migration here is the combined effect of the time allocation (the time a parent spends with their child) and information effect. The estimation results show that the net effect of parent's migration is not significant for most measures of children's health outcome.

It is surprising to see that the number of elderly in the household only has a few significant effects on children's health. The elderly in the household are likely to be children's grandparents. Intuitively, the care from grandparents could compensate the leave of children's parents. From the regression results, children who live with grandfathers take more calories. Children who live with their grandmothers do not have significantly better health outcome than those who do not live with their grandmothers. However the analysis in this paper focuses only on the measures of children's physical health, grandparents may have positive effects on children's mental health when children's parents are absent, which could be studied by future research.

The variances of the coefficients in the IV approach are obviously larger than the ones in the fixed effects model without correcting for the endogeneity. This is a sign that the instruments are not adding much variation. The variance is especially large for the dependent variable of child care. as the effective sample size is relatively small due to missing values for the child care variable. A total of 1048 observations were used to analyze the child care variable. There are 118 individuals that have more than one observation in the sample, among which 33 children's parents have changed their migration status.

Overall, the IV approach suggests that there were few significant causality effects of parents' migration on children's health outcomes. In contrast to the concern that left-behind children might suffer health problems without sufficient care from migrated parents, our empirical results show that the net effect of parents' migration on children's health is not necessarily negative. These results suggest that the effects of health information provided by migrated parents are important, and cannot be ignored.

4.7 Robustness Check

Of primary concern is that changes in household characteristics reflected in our data may be endogenous to children's health status. For instance, the changes in household income may be correlated with unobserved shocks that could also lead to changes in children's health. Moreover, household income may be correlated with migration decisions of household members. Such correlation may lead to biased estimates of migration. To rule out the possibility that the above results are driven by changes in endogenous household income, we estimated the regressions without including household income as a control variable. The effects of households' migration on children's health outcome and care are reported in Table 4.14a and Table 4.14b. The effects of fathers' migration on children's health outcomes and care are reported in Table 4.15a and Table 4.15b. The effects of mothers' migration on children's health outcomes and care are reported in Table 4.16a and Table 4.16b.

Compared to the previous estimates with household income as a control variable, the coefficients on migration are very similar in all regression analysis, with only small variations in the coefficients. The existence of the small variations might due to the coefficients of parents' migration also capturing the income effect from labor migration when we exclude the household income as a control variable.

Besides the household income, the number of elders in the household might be endogenous because this may be a factor in parents' migration decision-making. Therefore the estimates of the coefficients of migration might be biased. In order to address this issue, we estimated the effects of households' migration on child health outcomes and care from the fixed effects model without including the number of males/females over 60 in the household as control variables in Table 4.17a and Table 4.17b. The effects of fathers' migration on children's health outcomes and care in Table 4.18a and Table 4.18b, and the effects of mothers' migration on children's health outcomes and care in Table 4.19a and Table 4.19b.

The estimation results showed that the magnitude of the coefficients were very similar. It is worth noting that there were small changes in the standard deviation of some coefficients. One of the possible explanations is that the number of elders is positively correlated with migration decisions. As a consequence of multicollinearity, the variance is smaller in this robustness check. Albeit the change, the results are consistent with the previous findings.

4.8 Regression Results on Subsamples

Although the regression results show that there are few significant effect of parents' migration on children's health outcomes and care in general, parent's migration may have a significant effect on children in particular groups. In this section, we present the regression results from fixed effects model and IV approach on subsamples.

In total, we studied ten subgroups:[6] a) children who live in low income households, where low income is defined as household income level less than the average annual income level; b) children who live in high income households, where high income is defined as household income level higher than the average annual income level; c) children whose parents did not finish high school; d) children whose parents finished high school; e) children younger than age 5; f) children between

[6] I will add interactions instead of complete stratification for subgroup analysis in the future work.

ages 5 and 10; g) children who live with their grandparents; h) children who live in nuclear families; i) children who live in north China; j) children who live in south China. In addition, the regression results might be different in one-child families and families with more than one child. The income effect of parent migration is smaller for families with more than one child because the remittances spend for each child is less than those in one-child family. This smaller income effect might not be large enough to offset the negative effect caused by lacking of parenting time.[7]

Due to limitations of the data, some of the coefficients are not identifiable, particularly the coefficients of mother's migration, as the effective sample size is too small for some subsamples. The effective sample contains the children who have more than one observation in the data. Moreover the IVs are the county level average migration rate, not much variation was added by the IVs especially when the effective sample size was small. This problem is more serious for mother's migration because males migrate more often than females, and there is less variation in female's migration status than male's migration status. For the above reason, the regression results of mothers' migration are not reported here. The regression results for the subsample of children under age 5 and children with highly educated parents are not available for the same reason. A second problem is that due to the missing value problem, the effective sample size was too small in some subsamples for some variables to conduct fixed effect model analysis. For instance, child care data for several subsamples were not available.

Table (4.20a) and Table (4.20b) show the results of fixed effects model of household migration on children's health and care on subsamples using the IV approach. Table (4.21a) and Table (4.21b) show the results of fixed effects model of

[7]Regressions in the subgroups of one child families and families with more than one child will be added in future work.

fathers' migration.

Generally speaking, the IV approach shows that children between age 5 and 10 are significantly affected by fathers' migration. The effects are positive on children's calories and protein intake for children aged between 5 and 10 years. As I have mentioned, the effects of parents' migration can be both positive and negative. Positive effects include better access to nutritional information and products. Negative effects may include children not being in the care of either their mother or father. The results show that there are more positive effects of fathers' migration than negative effects for children aged between 5 and 10 years. No significant positive effects were found for other subsamples, possibly because children between the ages of 5 and 10 were in the midst of a crucial period of physical development. Most of the other coefficients were not significant due to the large standard deviations in the IV approach.

The regressions on subsamples show that parents' migration had significant effects on children's health outcome and care for children in particular groups. The results showed that the positive effects of parents' migration could offset the negative effects of parents' migration. Additionally, the positive effects outnumber the negative effects for children's nutrient intake in some subsamples.

4.9 CONCLUSION

In this paper, we studied left-behind children's health outcomes including height-for-age Z-score (HAZ), weight-for-age Z-score (WAZ), daily calorie intake, daily protein intake, the number of immunization shots received by children and whether children have been sick during the survey year. The evidence presented above showed that children with migrated parents did not necessarily have poorer health outcomes than children who lived with both parents. The robustness checks on the endogeneity assumption supported the findings that labor migration had no causal effect on the health of left-behind children. The fact that the results changed so little after excluding household income and the number of elders in the household suggests that parents' migration had no significant impact on children's health, that children's health is independent of household income and the number of elders in the household.

The regression results on subsamples showed that fathers' migration had significant positive effects on children's nutrient intake for children between 5 and 10 years of age. It showed that the positive effects of parents' migration out-number and could offset the negative effects of parents' migration. The regression results on subsamples provide some insights of the insignificance of the effects of parents' migration. The negative effects on children's health of parents' migration are possibly compensated by better access to nutrition information and products, the care from grandparents and the remittances that migrated parents are able to provide.

We have explored the possible mechanisms that may lead to better access to nutritional information. Future research should examine whether parental migration effects the social support that children receive and how children's health outcomes vary based on the duration of parents' migration. Nevertheless, these first steps into the investigation of this important topic cast further doubt on the view that those left-behind children in China always suffer from their parents' absence. These findings should encourage policy makers in areas of high migration to provide alternative sources of support for left-behind children.

Table 4.1: Parents Migration Rate for Children under age ten(CHNS)

	1997	2000	2004	2006	2009
Any Parent Migrated	0.06	0.10	0.17	0.21	0.14
Father Migrated Only	0.05	0.09	0.14	0.18	0.12
Mother Migrated Only	0.02	0.03	0.07	0.06	0.05
Both Parents Migrated	0.01	0.02	0.04	0.04	0.03
Number of Observations	927	785	614	585	531

Table 4.2a: Descriptive Statistics (CHNS)

Variables	Migrant household	Non-Migrant Household	t-stats of the difference
Weight (kg)	21.04	21.17	-0.32
	(0.36)[8]	(0.16)	(0.74)[9]
Height (cm)	114.90	113.94	0.98
	(0.89)	(0.37)	(0.33)
Weight-for-age Z-score	-0.49	-0.25	-3.41^{***}
	(0.07)	(0.03)	(0.00)
Height-for-age Z-score	-0.65	-0.50	$-1.80^{\cdot}$
	(0.08)	(0.03)	(0.07)
Calories (Kcal)	1362.68	1374.04	-0.34
	(30.00)	(13.87)	(0.73)
Protein (g)	40.65	42.75	$-1.90^{\cdot}$
	(1.00)	(0.48)	(0.06)
Calories/RDA	0.81	0.84	-1.26
	(0.02)	(0.01)	(0.21)
Protein/RDA	0.72	0.78	-3.03^{**}
	(0.02)	(0.01)	(0.00)
Number of immunization shots	6.53	8.97	$-1.88^{\cdot}$
	(1.15)	(0.60)	(0.06)
Whether the child has been cared by non-family member for the past week	0.39 (0.07)	0.47 (0.03)	-1.06 (0.29)
Num.obs	330	1871	2201

[1]standard deviation of the sample mean;
[2]p-value, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.2b: Descriptive Statistics (CHNS)

Variables	Migrant Father	Non-Migrant Father	t-stats
Weight (kg)	21.08	21.16	-0.18
	(0.39)[10]	(0.15)	(0.86)[11]
Height (cm)	115.04	113.95	1.04
	(0.97)	(0.37)	(0.30)
Weight-for-age Z-score	-0.48	-0.25	-2.86^{**}
	(0.07)	(0.03)	(0.00)
Height-for-age Z-score	-0.63	-0.51	-1.45
	(0.08)	(0.03)	(0.15)
Calories (Kcal)	1379.47	1371.43	0.21
	(33.35)	(13.63)	(0.83)
Protein (g)	41.22	42.62	-1.07
	(1.10)	(0.47)	(0.29)
Calories/RDA	0.82	0.83	-0.58
	(0.02)	(0.01)	(0.56)
Protein/RDA	0.73	0.77	$-1.90^{\cdot}$
	(0.02)	(0.01)	(0.06)
Number of immunization shots	6.42	8.92	-1.58
	(1.22)	(0.59)	(0.11)
Whether the child has been cared by non-family member for the past week	0.39	0.47	-0.83
	(0.08)	(0.03)	(0.41)
Num.obs	211	1990	2201

[1]standard deviation of the sample mean;
[2]p-value, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.2c: Descriptive Statistics (CHNS)

Variables	Migrant Mother	Non-Migrant Mother	t-stats
Weight (kg)	20.63	21.18	-0.87
	(0.57)[12]	(0.15)	(0.38)[13]
Height (cm)	113.73	114.10	-0.24
	(1.38)	(0.36)	(0.81)
Weight-for-age Z-score	-0.49	-0.27	$-1.95^{\cdot}$
	(0.11)	(0.03)	(0.05)
Height-for-age Z-score	-0.69	-0.52	-1.39
	(0.12)	(0.03)	(0.16)
Calories (Kcal)	1265.22	1378.60	-2.03^{*}
	(44.74)	(13.10)	(0.04)
Protein (g)	37.65	42.73	-2.67^{**}
	(1.52)	(0.45)	(0.01)
Calories/RDA	0.77	0.84	-2.16^{*}
	(0.02)	(0.01)	(0.03)
Protein/RDA	0.68	0.77	-2.73^{**}
	(0.03)	(0.01)	(0.01)
Number of immunization shots	5.52	8.79	-1.44
	(1.59)	(0.57)	(0.15)
Whether the child has been cared by non-family member for the past week	0.38	0.47	-0.68
	(0.06)	(0.03)	(0.50)
Num.obs	119	2082	2201

[1]standard deviation of the sample mean;
[2]p-value, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.3: Descriptive Statistics (CHNS) of Control Variables

Variables	Migrant household	Non-Migrant Household	t-stats
Household annual income (10000$)	2.30	2.82	-2.38^{*}
	(0.20)	(0.08)	(0.02)
Father's education	2.03	2.37	-6.75^{***}
	(0.04)	(0.02)	(0.00)
Mother's education	1.79	2.15	-6.34^{***}
	(0.05)	(0.03)	(0.00)
County level average	0.71	0.83	-3.45^{***}
income	(0.03)	(0.01)	(0.00)
County level average	56.72	58.85	-7.47^{***}
weight	(0.26)	(0.12)	(0.00)
County level average	158.48	160.12	-8.51^{***}
height	(0.18)	(0.08)	(0.00)
Number of male over	0.28	0.26	0.60
60 in the household	(0.03)	(0.01)	(0.55)
Child's gender	0.47	0.48	-0.06
(girls=1)	(0.03)	(0.01)	(0.95)
Number of female over	0.32	0.29	$1.00^{\cdot}$
60 in the household	(0.03)	(0.01)	(0.32)
Number of boys in	0.93	0.79	3.39^{***}
the household	(0.04)	(0.01)	(0.00)
Number of girls in	0.82	0.72	2.03^{*}
the household	(0.04)	(0.02)	(0.04)
County level average	1.02	1.00	1.16
calorie intake/RDA	(0.01)	(0.00)	(0.25)
County level average	1.30	1.30	0.04
protein intake/RDA	(0.02)	(0.01)	(0.97)
Children's age	6.22	5.94	2.006^{*}
	(0.13)	(0.05)	(0.05)
Father's age	32.60	34.16	-3.05^{**}
	(0.47)	(0.19)	(0.00)
Mother's age	31.25	32.32	-2.10^{*}
	(0.46)	(0.22)	(0.04)
Historical county level	0.26	0.15	9.50^{***}
male migration rate	(0.01)	(0.00)	(0.00)
Historical county level	0.16	0.09	8.58^{***}
female migration rate	(0.01)	(0.00)	(0.00)
Historical county level	0.32	0.20	10.00^{***}
household migration rate	(0.01)	(0.00)	(0.00)
Num.obs	330	1871	2201

[1]standard deviation;
[2]p-value $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$,
[3]historical county level migration rate: the average local migration rate from previous survey year.

Table 4.4a: OLS regression results: the effects of the household migration status

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Household migration status	0.07	0.13$^{\cdot}$	−2.95$^{\cdot}$	−0.03
	(0.06)	(0.07)	(1.51)	(0.04)
Household income	0.00	0.01	−0.35$^{\cdot}$	0.00
	(0.01)	(0.01)	(0.20)	(0.00)
Father education	0.07**	0.10**	−0.52	0.04*
	(0.03)	(0.03)	(0.68)	(0.02)
Mother education	0.02	0.06$^{\cdot}$	−0.92	0.01
	(0.03)	(0.03)	(0.69)	(0.02)
County average income	0.10*	0.24***	4.06*	−0.03
	(0.05)	(0.05)	(1.84)	(0.03)
County average weight	0.05***	0.03**	0.11	0.01*
	(0.01)	(0.01)	(0.22)	(0.01)
County average height	0.08***	0.08***	−0.42	−0.02$^{\cdot}$
	(0.01)	(0.01)	(0.32)	(0.01)
Male in household with age over 60	−0.06	−0.05	3.53*	−0.08*
	(0.06)	(0.07)	(1.42)	(0.04)
Female in household with age over 60	0.11*	0.10	0.52	−0.02
	(0.05)	(0.06)	(1.32)	(0.04)
Gender	−0.20**	−0.19*	−0.14	0.06
	(0.07)	(0.08)	(1.66)	(0.05)
Number of boys in household	−0.09$^{\cdot}$	−0.10$^{\cdot}$	0.78	−0.02
	(0.05)	(0.05)	(1.11)	(0.03)
Number of girls in household	−0.01	0.01	−0.88	−0.05$^{\cdot}$
	(0.04)	(0.05)	(1.07)	(0.03)
County average calorie consumption	−0.22	0.00	−8.67	0.30*
	(0.21)	(0.24)	(5.34)	(0.14)
County average protein consumption	0.14	0.24	3.95	−0.12
	(0.15)	(0.18)	(3.89)	(0.10)
Child age	−0.05***	0.03**	−0.10	0.02$^{\cdot}$
	(0.01)	(0.01)	(0.23)	(0.01)
Intercept	−15.46***	−16.22***	74.37$^{\cdot}$	1.68
	(1.67)	(1.92)	(41.52)	(1.09)
R^2	0.25	0.19	0.02	0.04
Adj. R^2	0.25	0.19	0.02	0.04
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.4b: OLS regression results: the effects of the household migration status

	Calorie	Protein	Calorie/RDA	Protein/RDA
Household Migration status	−30.15	−1.67	−0.02	−0.03
	(31.22)	(1.06)	(0.02)	(0.02)
Household income	−2.16	0.07	0.00	0.00
	(3.35)	(0.11)	(0.00)	(0.00)
Father education	29.90*	1.19**	0.02**	0.03**
	(13.12)	(0.45)	(0.01)	(0.01)
Mother education	28.58*	1.30**	0.02*	0.02**
	(13.17)	(0.45)	(0.01)	(0.01)
County average income	9.06	0.52	0.00	0.01
	(21.97)	(0.75)	(0.01)	(0.01)
County average weight	0.41	0.02	0.00	0.00
	(4.28)	(0.15)	(0.00)	(0.00)
County average height	0.99	0.08	0.00	0.00
	(6.24)	(0.21)	(0.00)	(0.00)
Male in household with age over 60	12.53	0.32	0.01	0.01
	(27.41)	(0.93)	(0.02)	(0.02)
Female in household with age over 60	−0.19	−0.20	0.00	−0.01
	(26.26)	(0.89)	(0.02)	(0.02)
Gender	−80.79*	−3.89***	−0.01	−0.04$^{\cdot}$
	(33.45)	(1.14)	(0.02)	(0.02)
Number of boys in household	3.52	−0.41	0.00	0.00
	(22.30)	(0.76)	(0.01)	(0.01)
Number of girls in household	−10.36	−0.23	−0.01	0.00
	(21.16)	(0.72)	(0.01)	(0.01)
County average calorie consumption	738.90***	−10.95**	0.42***	−0.22***
	(99.30)	(3.38)	(0.06)	(0.07)
County average protein consumption	195.62**	35.16***	0.13**	0.65***
	(73.58)	(2.51)	(0.05)	(0.05)
Child age	111.98***	3.21***	0.01***	0.01*
	(4.58)	(0.16)	(0.00)	(0.00)
Intercept	−566.10	−28.08	−0.29	−0.50
	(805.99)	(27.45)	(0.51)	(0.53)
R^2	0.29	0.30	0.11	0.18
Adj. R^2	0.29	0.30	0.11	0.18
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.5a: OLS regression results: the effects of the father's migration

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Father's migration status	0.07	$0.14^{\cdot}$	$-3.05^{\cdot}$	−0.03
	(0.07)	(0.08)	(1.60)	(0.05)
Household income	0.00	0.01	$-0.35^{\cdot}$	0.00
	(0.01)	(0.01)	(0.20)	(0.00)
Father education	0.07^{**}	0.10^{**}	−0.53	0.04^{*}
	(0.03)	(0.03)	(0.68)	(0.02)
Mother education	0.02	$0.05^{\cdot}$	−0.91	0.01
	(0.03)	(0.03)	(0.69)	(0.02)
County average income	0.10^{*}	0.24^{***}	4.09^{*}	−0.03
	(0.05)	(0.05)	(1.84)	(0.03)
County average weight	0.05^{***}	0.03^{**}	0.11	0.01^{*}
	(0.01)	(0.01)	(0.22)	(0.01)
County average height	0.08^{***}	0.08^{***}	−0.41	$-0.02^{\cdot}$
	(0.01)	(0.01)	(0.32)	(0.01)
Male in household with age over 60	−0.06	−0.06	3.58^{*}	-0.08^{*}
	(0.06)	(0.07)	(1.42)	(0.04)
Female in household with age over 60	0.11^{*}	0.10	0.48	−0.02
	(0.05)	(0.06)	(1.32)	(0.04)
Gender	-0.20^{**}	-0.19^{*}	−0.14	0.06
	(0.07)	(0.08)	(1.66)	(0.05)
Number of boys in household	$-0.09^{\cdot}$	$-0.10^{\cdot}$	0.81	−0.02
	(0.05)	(0.05)	(1.11)	(0.03)
Number of girls in household	−0.01	0.01	−0.85	$-0.05^{\cdot}$
	(0.04)	(0.05)	(1.07)	(0.03)
County average calorie consumption	−0.22	0.00	−8.55	0.30^{*}
	(0.21)	(0.24)	(5.34)	(0.14)
County average protein consumption	0.14	0.24	3.88	−0.12
	(0.15)	(0.18)	(3.89)	(0.10)
Child age	-0.05^{***}	0.03^{**}	−0.11	$0.02^{\cdot}$
	(0.01)	(0.01)	(0.23)	(0.01)
Intercept	-15.44^{***}	-16.18^{***}	$73.12^{\cdot}$	1.68
	(1.67)	(1.92)	(41.47)	(1.09)
R^2	0.25	0.19	0.02	0.04
Adj. R^2	0.25	0.19	0.02	0.04
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.5b: OLS regression results: the effects of the father's migration

	Calorie	Protein	Calorie/RDA	Protein/RDA
Father's Migration status	−12.72	−1.05	−0.01	−0.02
	(33.16)	(1.13)	(0.02)	(0.02)
Household income	−2.12	0.07	0.00	0.00
	(3.35)	(0.11)	(0.00)	(0.00)
Father education	30.27^{*}	1.20^{**}	0.02^{**}	0.03^{**}
	(13.13)	(0.45)	(0.01)	(0.01)
Mother education	28.74^{*}	1.31^{**}	0.02^{*}	0.02^{**}
	(13.17)	(0.45)	(0.01)	(0.01)
County average income	9.00	0.52	0.00	0.01
	(21.98)	(0.75)	(0.01)	(0.01)
County average weight	0.41	0.02	0.00	0.00
	(4.29)	(0.15)	(0.00)	(0.00)
County average height	1.30	0.09	0.00	0.00
	(6.24)	(0.21)	(0.00)	(0.00)
Male in household with age over 60	12.82	0.34	0.01	0.01
	(27.42)	(0.93)	(0.02)	(0.02)
Female in household with age over 60	−0.31	−0.21	0.00	−0.01
	(26.26)	(0.89)	(0.02)	(0.02)
Gender	-80.83^{*}	-3.89^{***}	−0.01	$-0.04^{\cdot}$
	(33.46)	(1.14)	(0.02)	(0.02)
Number of boys in household	3.06	−0.42	0.00	0.00
	(22.32)	(0.76)	(0.01)	(0.01)
Number of girls in household	−10.66	−0.24	−0.01	0.00
	(21.17)	(0.72)	(0.01)	(0.01)
County average calorie consumption	740.12^{***}	-10.90^{**}	0.42^{***}	-0.22^{***}
	(99.32)	(3.38)	(0.06)	(0.07)
County average protein consumption	194.04^{**}	35.10^{***}	0.13^{**}	0.65^{***}
	(73.58)	(2.51)	(0.05)	(0.05)
Child age	111.88^{***}	3.20^{***}	0.01^{***}	0.01^{*}
	(4.58)	(0.16)	(0.00)	(0.00)
Intercept	−617.76	−30.21	−0.31	−0.53
	(805.08)	(27.42)	(0.51)	(0.53)
R^2	0.29	0.30	0.11	0.18
Adj. R^2	0.29	0.30	0.11	0.18
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.6a: OLS regression results: the effects of the mother's migration

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Mother's migration status	$0.18^{\cdot}$	$0.20^{\cdot}$	-3.47	0.02
	(0.10)	(0.11)	(2.28)	(0.06)
Household income	0.00	0.01	$-0.33^{\cdot}$	0.00
	(0.01)	(0.01)	(0.20)	(0.00)
Father education	0.07^{**}	0.10^{**}	-0.44	0.04^{*}
	(0.03)	(0.03)	(0.68)	(0.02)
Mother education	0.02	$0.06^{\cdot}$	-0.94	0.01
	(0.03)	(0.03)	(0.69)	(0.02)
County average income	0.10^{*}	0.24^{***}	4.00^{*}	-0.03
	(0.05)	(0.05)	(1.84)	(0.03)
County average weight	0.05^{***}	0.03^{**}	0.09	0.01^{*}
	(0.01)	(0.01)	(0.22)	(0.01)
County average height	0.08^{***}	0.08^{***}	-0.39	$-0.01^{\cdot}$
	(0.01)	(0.01)	(0.32)	(0.01)
Male in household with age over 60	-0.06	-0.05	3.51^{*}	-0.08^{*}
	(0.06)	(0.07)	(1.42)	(0.04)
Female in household with age over 60	0.11^{*}	0.10	0.59	-0.02
	(0.05)	(0.06)	(1.32)	(0.04)
Gender	-0.20^{**}	-0.19^{*}	-0.19	0.06
	(0.07)	(0.08)	(1.66)	(0.05)
Number of boys in household	$-0.08^{\cdot}$	$-0.09^{\cdot}$	0.63	-0.02
	(0.05)	(0.05)	(1.11)	(0.03)
Number of girls in household	0.00	0.02	-0.98	$-0.05^{\cdot}$
	(0.04)	(0.05)	(1.07)	(0.03)
County average calorie consumption	-0.23	-0.02	-8.33	0.30^{*}
	(0.21)	(0.24)	(5.34)	(0.14)
County average protein consumption	0.14	0.25	3.69	-0.13
	(0.15)	(0.18)	(3.88)	(0.10)
Child age	-0.05^{***}	0.03^{**}	-0.12	$0.02^{\cdot}$
	(0.01)	(0.01)	(0.23)	(0.01)
Intercept	-15.56^{***}	-16.16^{***}	$70.76^{\cdot}$	1.58
	(1.67)	(1.92)	(41.44)	(1.08)
R^2	0.25	0.19	0.02	0.04
Adj. R^2	0.25	0.19	0.02	0.04
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.6b: OLS regression results: the effects of the mother's migration

	Calorie	Protein	Calorie/RDA	Protein/RDA
Mother's Migration status	-116.12^{*}	-3.45^{*}	-0.06^{*}	$-0.06^{\cdot}$
	(47.71)	(1.63)	(0.03)	(0.03)
Household income	-2.19	0.07	0.00	0.00
	(3.34)	(0.11)	(0.00)	(0.00)
Father education	30.43^{*}	1.22^{**}	0.02^{**}	0.03^{**}
	(13.08)	(0.45)	(0.01)	(0.01)
Mother education	27.60^{*}	1.28^{**}	0.02^{*}	0.02^{**}
	(13.16)	(0.45)	(0.01)	(0.01)
County average income	9.35	0.53	0.00	0.01
	(21.95)	(0.75)	(0.01)	(0.01)
County average weight	0.23	0.01	0.00	0.00
	(4.28)	(0.15)	(0.00)	(0.00)
County average height	0.50	0.07	0.00	0.00
	(6.23)	(0.21)	(0.00)	(0.00)
Male in household with age over 60	11.15	0.28	0.01	0.01
	(27.39)	(0.93)	(0.02)	(0.02)
Female in household with age over 60	0.52	-0.18	0.00	-0.01
	(26.23)	(0.89)	(0.02)	(0.02)
Gender	-81.57^{*}	-3.91^{***}	-0.01	$-0.04^{\cdot}$
	(33.42)	(1.14)	(0.02)	(0.02)
Number of boys in household	1.49	-0.50	0.00	-0.01
	(22.25)	(0.76)	(0.01)	(0.01)
Number of girls in household	-12.25	-0.31	-0.01	-0.01
	(21.13)	(0.72)	(0.01)	(0.01)
County average calorie consumption	744.00^{***}	-10.74^{**}	0.43^{***}	-0.22^{***}
	(99.18)	(3.38)	(0.06)	(0.07)
County average protein consumption	193.20^{**}	35.02^{***}	0.13^{**}	0.65^{***}
	(73.44)	(2.50)	(0.05)	(0.05)
Child age	111.85^{***}	3.20^{***}	0.01^{***}	0.01^{*}
	(4.57)	(0.16)	(0.00)	(0.00)
Intercept	-471.22	-27.27	-0.23	-0.49
	(804.12)	(27.40)	(0.51)	(0.53)
R^2	0.29	0.30	0.11	0.18
Adj. R^2	0.29	0.30	0.11	0.18
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.7a: Fixed effects model results of the effects of the household migration status on children's health outcome and care

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Household migration status	$-0.20^{\cdot}$	-0.13	-5.11	-0.02
	(0.10)	(0.12)	(3.78)	(0.13)
Household income	-0.03^{*}	-0.02	-0.28	-0.01
	(0.01)	(0.01)	(0.57)	(0.01)
County average income	0.27^{**}	0.08	3.84	0.05
	(0.09)	(0.10)	(5.60)	(0.10)
County average weight	0.03	-0.04	-0.10	0.02
	(0.03)	(0.03)	(1.13)	(0.04)
County average height	0.03	0.05	-1.65	-0.01
	(0.03)	(0.04)	(1.21)	(0.04)
Male in household with age over 60	-0.07	0.08	0.75	0.06
	(0.16)	(0.18)	(6.52)	(0.23)
Female in household with age over 60	0.13	0.30	11.13	0.10
	(0.19)	(0.21)	(7.50)	(0.28)
Number of boys in household	-0.04	-0.17	9.50^{*}	0.07
	(0.12)	(0.13)	(4.12)	(0.17)
Number of girls in household	-0.08	-0.21	-0.78	-0.03
	(0.12)	(0.14)	(5.21)	(0.18)
County average calorie consumption	0.28	0.14	1.27	0.87
	(0.35)	(0.39)	(14.05)	(0.54)
County average protein consumption	0.08	0.29	-4.57	-0.36
	(0.25)	(0.29)	(9.00)	(0.38)
Child age	-0.04^{*}	0.09^{***}	1.35^{*}	0.00
	(0.02)	(0.02)	(0.62)	(0.03)
R^2	0.04	0.10	0.07	0.03
Adj. R^2	0.01	0.03	0.02	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.7b: Fixed effects model results of the effects of the household migration status on children's health outcome and care

	Calorie	Protein	Calorie/RDA	Protein/RDA
Household Migration status	-25.15	1.20	-0.01	0.03
	(69.77)	(2.26)	(0.05)	(0.05)
Household income	-3.97	-0.24	0.00	-0.01
	(7.34)	(0.24)	(0.00)	(0.00)
County average income	-28.54	-0.04	-0.02	-0.01
	(60.74)	(1.97)	(0.04)	(0.04)
County average weight	7.94	0.29	0.01	0.01
	(20.05)	(0.65)	(0.01)	(0.01)
County average height	12.92	$1.15^{\cdot}$	0.01	$0.02^{\cdot}$
	(20.71)	(0.67)	(0.01)	(0.01)
Male in household with age over 60	$207.50^{\cdot}$	3.46	$0.13^{\cdot}$	0.07
	(107.41)	(3.48)	(0.07)	(0.07)
Female in household with age over 60	48.21	-1.87	0.02	-0.08
	(125.27)	(4.06)	(0.09)	(0.08)
Number of boys in household	43.85	-0.55	0.02	-0.02
	(78.20)	(2.54)	(0.05)	(0.05)
Number of girls in household	27.00	-0.89	0.02	-0.03
	(82.23)	(2.67)	(0.06)	(0.06)
County average calorie consumption	1159.18^{***}	4.22	0.65^{***}	0.02
	(232.22)	(7.53)	(0.16)	(0.16)
County average protein consumption	47.99	26.66^{***}	0.07	0.54^{***}
	(170.03)	(5.51)	(0.12)	(0.12)
Child age	97.59^{***}	2.56^{***}	0.00	0.00
	(11.05)	(0.36)	(0.01)	(0.01)
R^2	0.27	0.23	0.09	0.10
Adj. R^2	0.07	0.06	0.02	0.03
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.8a: Fixed effects model results of the effects of the father's migration status on children's health outcome and care

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Father migration status	$-0.19^{\cdot}$	-0.20	-4.19	-0.06
	(0.11)	(0.12)	(3.80)	(0.14)
Household income	-0.02^{*}	-0.02	-0.27	-0.01
	(0.01)	(0.01)	(0.57)	(0.01)
County average income	0.27^{**}	0.07	3.90	0.05
	(0.09)	(0.10)	(5.60)	(0.10)
County average weight	0.03	-0.04	-0.07	0.02
	(0.03)	(0.03)	(1.13)	(0.04)
County average height	0.03	0.04	-1.60	-0.01
	(0.03)	(0.03)	(1.21)	(0.04)
Male in household with age over 60	-0.06	0.09	1.02	0.06
	(0.16)	(0.18)	(6.52)	(0.23)
Female in household with age over 60	0.12	0.30	11.00	0.10
	(0.19)	(0.21)	(7.51)	(0.28)
Number of boys in household	-0.03	-0.16	9.68^{*}	0.07
	(0.12)	(0.13)	(4.12)	(0.17)
Number of girls in household	-0.08	-0.21	-0.84	-0.02
	(0.12)	(0.14)	(5.21)	(0.18)
County average calorie consumption	0.27	0.13	0.77	0.87
	(0.35)	(0.39)	(14.07)	(0.54)
County average protein consumption	0.08	0.29	-4.36	-0.36
	(0.25)	(0.29)	(9.01)	(0.38)
Child age	-0.04^{*}	0.09^{***}	1.31^{*}	0.01
	(0.02)	(0.02)	(0.62)	(0.03)
R^2	0.04	0.11	0.07	0.03
Adj. R^2	0.01	0.03	0.02	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.8b: Fixed effects model results of the effects of the father's migration status on children's health outcome and care

	Calorie	Protein	Calorie/RDA	Protein/RDA
Father Migration status	7.54	3.18	0.01	0.07
	(72.23)	(2.34)	(0.05)	(0.05)
Household income	-3.80	-0.24	0.00	-0.01
	(7.33)	(0.24)	(0.00)	(0.00)
County average income	-28.19	0.11	-0.02	-0.01
	(60.84)	(1.97)	(0.04)	(0.04)
County average weight	8.14	0.31	0.01	0.01
	(20.05)	(0.65)	(0.01)	(0.01)
County average height	13.89	$1.19^{\cdot}$	0.01	$0.02^{\cdot}$
	(20.67)	(0.67)	(0.01)	(0.01)
Male in household with age over 60	$208.81^{\cdot}$	3.35	$0.13^{\cdot}$	0.07
	(107.36)	(3.48)	(0.07)	(0.07)
Female in household with age over 60	46.34	-1.87	0.02	-0.08
	(125.21)	(4.05)	(0.09)	(0.08)
Number of boys in household	44.25	-0.63	0.02	-0.02
	(78.21)	(2.53)	(0.05)	(0.05)
Number of girls in household	24.85	-1.03	0.02	-0.03
	(82.27)	(2.66)	(0.06)	(0.06)
County average calorie consumption	1161.00^{***}	4.44	0.65^{***}	0.03
	(232.30)	(7.52)	(0.16)	(0.16)
County average protein consumption	46.13	26.49^{***}	0.07	0.54^{***}
	(170.09)	(5.51)	(0.12)	(0.12)
Child age	96.94^{***}	2.52^{***}	0.00	0.00
	(11.08)	(0.36)	(0.01)	(0.01)
R^2	0.27	0.23	0.09	0.10
Adj. R^2	0.07	0.06	0.02	0.03
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.9a: Fixed effects model results of the effects of the mother's migration status on children's health outcome and care

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Mother migration status	0.00	−0.10	−4.31	0.23
	(0.15)	(0.17)	(5.85)	(0.22)
Household income	−0.02*	−0.02	−0.25	−0.01
	(0.01)	(0.01)	(0.57)	(0.01)
County average income	0.27**	0.08	3.96	0.06
	(0.09)	(0.10)	(5.61)	(0.10)
County average weight	0.03	−0.04	0.01	0.02
	(0.03)	(0.03)	(1.12)	(0.04)
County average height	0.03	0.05	−1.57	−0.01
	(0.03)	(0.03)	(1.21)	(0.04)
Male in household with age over 60	−0.06	0.08	0.85	0.06
	(0.16)	(0.18)	(6.54)	(0.23)
Female in household with age over 60	0.12	0.30	10.94	0.13
	(0.19)	(0.21)	(7.52)	(0.28)
Number of boys in household	−0.03	−0.18	9.03*	0.09
	(0.12)	(0.13)	(4.20)	(0.18)
Number of girls in household	−0.10	−0.22	−0.97	−0.03
	(0.12)	(0.14)	(5.21)	(0.18)
County average calorie consumption	0.29	0.14	1.53	0.92$^{\cdot}$
	(0.35)	(0.39)	(14.09)	(0.54)
County average protein consumption	0.07	0.29	−4.57	−0.43
	(0.25)	(0.29)	(9.02)	(0.38)
Child age	−0.04*	0.09***	1.23*	0.00
	(0.02)	(0.02)	(0.61)	(0.03)
R^2	0.04	0.10	0.06	0.04
Adj. R^2	0.01	0.03	0.02	0.01
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.9b: Fixed effects model results of the effects of the mother's migration status on children's health outcome and care

	Calorie	Protein	Calorie/RDA	Protein/RDA
Mother Migration status	−113.18	−1.52	−0.06	0.00
	(102.70)	(3.33)	(0.07)	(0.07)
Household income	−4.11	−0.25	0.00	−0.01
	(7.32)	(0.24)	(0.00)	(0.00)
County average income	−27.98	−0.03	−0.02	−0.01
	(60.68)	(1.97)	(0.04)	(0.04)
County average weight	8.15	0.28	0.01	0.01
	(20.03)	(0.65)	(0.01)	(0.01)
County average height	12.55	1.10	0.01	0.02
	(20.60)	(0.67)	(0.01)	(0.01)
Male in household with age over 60	198.61$^{\cdot}$	3.26	0.12$^{\cdot}$	0.07
	(107.65)	(3.49)	(0.07)	(0.07)
Female in household with age over 60	56.99	−1.65	0.03	−0.08
	(125.43)	(4.07)	(0.09)	(0.09)
Number of boys in household	30.20	−0.76	0.01	−0.02
	(79.17)	(2.57)	(0.05)	(0.05)
Number of girls in household	22.48	−0.85	0.02	−0.02
	(82.07)	(2.66)	(0.06)	(0.06)
County average calorie consumption	1155.05***	4.09	0.65***	0.02
	(232.03)	(7.53)	(0.16)	(0.16)
County average protein consumption	55.93	26.84***	0.08	0.54***
	(170.04)	(5.52)	(0.12)	(0.12)
Child age	97.88***	2.60***	0.00	0.00
	(10.98)	(0.36)	(0.01)	(0.01)
R^2	0.27	0.23	0.09	0.10
Adj. R^2	0.07	0.06	0.02	0.03
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.10: First Stage fixed effects Regression Results

	Father's migration status	Mother's migration status	Household migration status
County level male migration rate	-0.3185^{*}		
	(0.1236)		
County level female migration rate		-0.2548^{*}	
		(0.1131)	
County level household migration rate			-0.3266^{**}
			(0.1212)
Father's age	-0.1506^{*}		-0.1538^{*}
	(0.0667)		(0.0692)
Mother's age		0.0132	0.0260
		(0.0577)	(0.0842)
Household income	-0.0031	-0.0028	-0.0048
	(0.0041)	(0.0030)	(0.0043)
Male in household with age over 60	0.0214	-0.0854^{*}	-0.0472
	(0.0606)	(0.0431)	(0.0628)
Female in household with age over 60	0.0038	$0.0902^{\cdot}$	0.0456
	(0.0702)	(0.0499)	(0.0727)
Number of children in the family	0.0001	-0.0132	-0.0050
	(0.0135)	(0.0096)	(0.0140)
County average income	-0.0353	0.0134	0.0092
	(0.0344)	(0.0248)	(0.0363)
Children's age	0.1724^{**}	-0.0067	0.1475
	(0.0665)	(0.0581)	(0.1086)
R^2	0.0381	0.0264	0.0373
Adj. R^2	0.0103	0.0071	0.0100
Num. obs.	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.11a: Fixed effects model results of the effects of the household migration status on children's health outcome and care: IV approach

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Household Migration status	2.07	0.19	26.62	0.08
	(1.35)	(1.14)	(39.20)	(18.46)
Household income	-0.01	-0.02	-0.08	-0.01
	(0.02)	(0.01)	(0.67)	(0.16)
County average income	0.27^{*}	0.08	5.88	0.05
	(0.12)	(0.10)	(6.61)	(0.61)
County average weight	0.05	-0.03	0.98	0.02
	(0.04)	(0.03)	(1.81)	(0.36)
County average height	$0.10^{\cdot}$	0.06	-0.91	-0.01
	(0.06)	(0.05)	(1.60)	(0.30)
Male in household with age over 60	0.05	0.10	3.47	0.05
	(0.23)	(0.19)	(7.88)	(0.65)
Female in household with age over 60	-0.02	0.28	7.55	0.11
	(0.27)	(0.23)	(9.31)	(0.54)
Number of boys in household	0.01	-0.16	10.28^{*}	0.06
	(0.16)	(0.14)	(4.60)	(1.24)
Number of girls in household	-0.23	-0.23	-2.53	-0.04
	(0.19)	(0.16)	(6.08)	(2.15)
County average calorie consumption	0.38	0.16	0.24	0.88
	(0.47)	(0.40)	(15.41)	(1.19)
County average protein consumption	-0.04	0.27	-6.11	-0.37
	(0.35)	(0.30)	(10.02)	(1.53)
Child age	-0.08^{*}	0.08^{**}	0.27	0.00
	(0.03)	(0.03)	(1.49)	(0.39)
R^2	0.00	0.09	0.01	0.03
Adj. R^2	0.00	0.03	0.00	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.11b: Fixed effects model results of the effects of the household migration status on children's health outcome and care: IV approach

	Calorie	Protein	Calorie/RDA	Protein/RDA
Household Migration status	1060.38	20.23	0.49	0.20
	(799.13)	(23.09)	(0.50)	(0.46)
Household income	2.16	−0.14	0.00	−0.01
	(9.80)	(0.28)	(0.01)	(0.01)
County average income	−28.57	−0.04	−0.02	−0.01
	(72.15)	(2.08)	(0.05)	(0.04)
County average weight	13.89	0.39	0.01	0.01
	(24.21)	(0.70)	(0.02)	(0.01)
County average height	46.75	$1.74^{\cdot}$	0.03	0.03
	(34.91)	(1.01)	(0.02)	(0.02)
Male in household with age over 60	268.37^{*}	4.53	$0.16^{\cdot}$	0.08
	(135.15)	(3.90)	(0.08)	(0.08)
Female in household with age over 60	−24.43	−3.14	−0.01	−0.09
	(158.03)	(4.57)	(0.10)	(0.09)
Number of boys in household	67.18	−0.14	0.03	−0.01
	(94.45)	(2.73)	(0.06)	(0.05)
Number of girls in household	−42.95	−2.11	−0.01	−0.04
	(110.29)	(3.19)	(0.07)	(0.06)
County average calorie consumption	1209.77^{***}	5.11	0.67^{***}	0.03
	(278.32)	(8.04)	(0.17)	(0.16)
County average protein consumption	−8.44	25.67^{***}	0.05	0.53^{***}
	(206.16)	(5.96)	(0.13)	(0.12)
Child age	76.80^{***}	2.20^{***}	−0.01	−0.01
	(20.10)	(0.58)	(0.01)	(0.01)
R^2	0.12	0.17	0.03	0.08
Adj. R^2	0.03	0.04	0.01	0.02
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.12a: Fixed effects model results of the effects of the father's migration status on children's health outcome and care: IV approach

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Father Migration status	2.12	0.87	37.40	−14.95
	(1.38)	(1.25)	(41.49)	(229.20)
Household income	−0.02	−0.01	−0.05	−0.16
	(0.02)	(0.01)	(0.69)	(2.32)
County average income	0.37^{**}	0.12	6.59	−0.23
	(0.14)	(0.12)	(7.00)	(4.43)
County average weight	0.05	−0.03	1.30	0.19
	(0.04)	(0.04)	(1.88)	(2.75)
County average height	0.09	0.07	−0.89	−0.28
	(0.05)	(0.05)	(1.56)	(4.12)
Male in household with age over 60	−0.09	0.07	2.72	0.45
	(0.21)	(0.19)	(7.71)	(6.34)
Female in household with age over 60	0.07	0.27	6.59	−0.41
	(0.25)	(0.23)	(9.70)	(8.33)
Number of boys in household	−0.08	−0.18	$9.10^{\cdot}$	0.98
	(0.16)	(0.14)	(4.79)	(14.10)
Number of girls in household	−0.24	−0.28	−3.02	1.75
	(0.19)	(0.17)	(6.39)	(27.34)
County average calorie consumption	0.47	0.22	4.03	0.02
	(0.48)	(0.43)	(16.55)	(14.02)
County average protein consumption	−0.09	0.21	−8.85	−0.13
	(0.35)	(0.32)	(11.31)	(5.04)
Child age	-0.09^{*}	0.07^{*}	−0.04	0.35
	(0.04)	(0.03)	(1.52)	(5.37)
R^2	0.00	0.04	0.00	0.00
Adj. R^2	0.00	0.01	0.00	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.12b: Fixed effects model results of the effects of the father's migration status on children's health outcome and care: IV approach

	Calorie	Protein	Calorie/RDA	Protein/RDA
Father Migration status	943.30	14.45	0.40	−0.03
	(788.91)	(22.98)	(0.50)	(0.47)
Household income	−0.29	−0.20	0.00	−0.01
	(8.82)	(0.26)	(0.01)	(0.01)
County average income	14.92	0.63	0.00	−0.01
	(77.86)	(2.27)	(0.05)	(0.05)
County average weight	15.98	0.40	0.01	0.01
	(23.66)	(0.69)	(0.01)	(0.01)
County average height	36.80	1.47$^{\cdot}$	0.02	0.02
	(30.31)	(0.88)	(0.02)	(0.02)
Male in household with age over 60	195.99	3.20	0.12	0.07
	(122.17)	(3.56)	(0.08)	(0.07)
Female in household with age over 60	23.31	−2.14	0.01	−0.08
	(143.24)	(4.17)	(0.09)	(0.09)
Number of boys in household	26.73	−0.84	0.01	−0.02
	(89.86)	(2.62)	(0.06)	(0.05)
Number of girls in household	−40.25	−1.81	−0.01	−0.02
	(108.06)	(3.15)	(0.07)	(0.06)
County average calorie consumption	1242.30***	5.42	0.68***	0.02
	(272.01)	(7.92)	(0.17)	(0.16)
County average protein consumption	−22.23	25.66***	0.04	0.55***
	(201.15)	(5.86)	(0.13)	(0.12)
Child age	76.80***	2.28***	0.00	0.00
	(21.05)	(0.61)	(0.01)	(0.01)
R^2	0.15	0.21	0.04	0.10
Adj. R^2	0.04	0.06	0.01	0.03
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.13a: Fixed effects model results of the effects of the mother's migration status on children's health outcome and care: IV approach

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Mother Migration status	2.21	−0.58	21.18	1.76
	(2.11)	(2.07)	(47.55)	(4.04)
Household income	−0.02	−0.02	−0.26	−0.01
	(0.01)	(0.01)	(0.58)	(0.01)
County average income	0.26*	0.09	5.20	0.11
	(0.11)	(0.10)	(6.20)	(0.17)
County average weight	0.03	−0.04	0.35	0.03
	(0.03)	(0.03)	(1.31)	(0.06)
County average height	0.06	0.04	−1.33	0.00
	(0.04)	(0.04)	(1.32)	(0.05)
Male in household with age over 60	0.14	0.03	2.87	0.10
	(0.27)	(0.26)	(7.68)	(0.29)
Female in household with age over 60	−0.08	0.34	8.66	0.31
	(0.29)	(0.29)	(8.79)	(0.56)
Number of boys in household	0.24	−0.24	12.53	0.27
	(0.30)	(0.29)	(7.78)	(0.52)
Number of girls in household	−0.04	−0.23	−1.51	−0.04
	(0.15)	(0.15)	(5.44)	(0.21)
County average calorie consumption	0.39	0.12	−1.02	1.26
	(0.42)	(0.41)	(15.21)	(1.08)
County average protein consumption	−0.11	0.33	−6.01	−0.87
	(0.34)	(0.34)	(9.63)	(1.24)
Child age	−0.06*	0.09***	0.93	−0.03
	(0.02)	(0.02)	(0.84)	(0.07)
R^2	0.00	0.09	0.03	0.02
Adj. R^2	0.00	0.03	0.01	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.13b: Fixed effects model results of the effects of the mother's migration status on children's health outcome and care: IV approach

	Calorie	Protein	Calorie/RDA	Protein/RDA
Mother Migration status	1742.55	26.08	0.95	0.47
	(1514.51)	(43.02)	(0.96)	(0.86)
Household income	0.49	−0.19	0.00	−0.01
	(9.87)	(0.27)	(0.01)	(0.01)
County average income	−37.09	−0.17	−0.02	−0.01
	(76.02)	(2.09)	(0.05)	(0.04)
County average weight	6.85	0.26	0.01	0.01
	(24.99)	(0.69)	(0.02)	(0.01)
County average height	31.53	1.38˙	0.02	0.03
	(29.97)	(0.83)	(0.02)	(0.02)
Male in household with age over 60	367.56˙	5.77	0.21˙	0.11
	(192.07)	(5.37)	(0.12)	(0.11)
Female in household with age over 60	−114.50	−4.20	−0.07	−0.12
	(209.53)	(5.85)	(0.13)	(0.12)
Number of boys in household	262.89	2.70	0.14	0.04
	(213.42)	(6.02)	(0.14)	(0.12)
Number of girls in household	70.07	−0.14	0.04	−0.01
	(109.40)	(3.02)	(0.07)	(0.06)
County average calorie consumption	1241.91***	5.39	0.69***	0.04
	(297.79)	(8.21)	(0.19)	(0.17)
County average protein consumption	−95.71	24.59***	0.00	0.51***
	(245.26)	(6.80)	(0.16)	(0.14)
Child age	85.21***	2.41***	0.00	−0.01
	(17.13)	(0.48)	(0.01)	(0.01)
R^2	0.09	0.16	0.01	0.06
Adj. R^2	0.03	0.04	0.00	0.02
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.14a: Robustness Check 1: the effects of the household migration status on children's health outcome and care without household income as a control variable

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Household Migration status	2.09	0.22	26.66	0.48
	(1.34)	(1.13)	(39.02)	(12.48)
R^2	0.00	0.09	0.01	0.00
Adj. R^2	0.00	0.02	0.00	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.14b: Robustness Check 1: the effects of the household migration status on children's health outcome and care without household income as a control variable

	Calorie	Protein	Calorie/RDA	Protein/RDA
Household Migration status	1056.39	20.48	0.50	0.21
	(789.50)	(22.87)	(0.50)	(0.46)
R^2	0.12	0.16	0.03	0.08
Adj. R^2	0.03	0.04	0.01	0.02
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.15a: Robustness Check 1: the effects of the father's migration status on children's health outcome and care without household income as a control variable

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Father Migration status	2.13	0.89	37.42	−8.80
	(1.38)	(1.25)	(41.37)	(83.99)
R^2	0.00	0.04	0.00	0.00
Adj. R^2	0.00	0.01	0.00	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.15b: Robustness Check 1: the effects of the father's migration status on children's health outcome and care without household income as a control variable

	Calorie	Protein	Calorie/RDA	Protein/RDA
Father Migration status	943.60	14.66	0.41	−0.02
	(785.27)	(22.90)	(0.50)	(0.47)
R^2	0.15	0.21	0.04	0.09
Adj. R^2	0.04	0.06	0.01	0.02
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.16a: Robustness Check 1: the effects of the mother's migration status on children's health outcome and care without household income as a control variable

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Mother Migration status	2.11	−0.67	20.95	2.22
	(2.12)	(2.11)	(47.46)	(4.57)
R^2	0.00	0.09	0.03	0.01
Adj. R^2	0.00	0.02	0.01	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.16b: Robustness Check 1: the effects of the mother's migration status on children's health outcome and care without household income as a control variable

	Calorie	Protein	Calorie/RDA	Protein/RDA
Mother Migration status	1745.04	26.43	0.94	0.44
	(1533.75)	(42.86)	(0.97)	(0.86)
R^2	0.09	0.16	0.01	0.06
Adj. R^2	0.03	0.04	0.00	0.02
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.17a: Robustness Check 2: the effects of the household migration status on children's health outcome and care without the number of elders as control variables

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Household Migration status	2.07	0.19	28.87	0.13
	(1.35)	(1.15)	(39.34)	(20.97)
R^2	0.00	0.09	0.00	0.02
Adj. R^2	0.00	0.02	0.00	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.17b: Robustness Check 2: the effects of the household migration status on children's health outcome and care without the number of elders as control variables

	Calorie	Protein	Calorie/RDA	Protein/RDA
Household Migration status	1064.49	20.31	0.50	0.20
	(802.52)	(23.11)	(0.50)	(0.46)
R^2	0.12	0.16	0.02	0.08
Adj. R^2	0.03	0.04	0.01	0.02
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.18a: Robustness Check 2: the effects of the father's migration status on children's health outcome and care without the number of elders as control variables

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Father Migration status	2.11	0.85	38.06	−14.87
	(1.38)	(1.25)	(41.67)	(225.42)
R^2	0.00	0.04	0.00	0.00
Adj. R^2	0.00	0.01	0.00	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.18b: Robustness Check 2: the effects of the father's migration status on children's health outcome and care without the number of elders as control variables

	Calorie	Protein	Calorie/RDA	Protein/RDA
Father Migration status	940.20	14.63	0.40	−0.02
	(790.83)	(23.02)	(0.50)	(0.47)
R^2	0.15	0.21	0.04	0.09
Adj. R^2	0.04	0.06	0.01	0.03
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.19a: Robustness Check 2: the effects of the mother's migration status on children's health outcome and care without the number of elders as control variables

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Mother Migration status	2.20	−0.55	23.11	2.30
	(2.10)	(2.06)	(46.78)	(5.39)
R^2	0.00	0.09	0.03	0.01
Adj. R^2	0.00	0.02	0.01	0.00
Num. obs.	2201	2201	1491	1048

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.19b: Robustness Check 2: the effects of the mother's migration status on children's health outcome and care without the number of elders as control variables

	Calorie	Protein	Calorie/RDA	Protein/RDA
Mother Migration status	1741.51	26.72	0.95	0.47
	(1513.70)	(43.79)	(0.96)	(0.85)
R^2	0.09	0.16	0.01	0.05
Adj. R^2	0.02	0.04	0.00	0.01
Num. obs.	2201	2201	2201	2201

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.20a: Fixed effects model results of the effects of the household migration status on children's health outcome and care on subsamples: IV approach

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Household Migration status	3.39	0.08	75.25	−0.44
(Low income household)	(3.59)	(2.37)	(133.99)	(1.91)
Household Migration status	3.25	−4.97	*N.A.*[14]	*N.A.*
(High income household)	(8.35)	(9.78)	*N.A.*	*N.A.*
Household Migration status	2.45	1.68	65.05	*N.A.*
(Parents with low education level)	(1.57)	(1.42)	(71.59)	*N.A.*
Household Migration status	1.49	0.48	5.30	*N.A.*
(Child above age 5)	(0.98)	(0.92)	(21.14)	*N.A.*
Household Migration status	4.08	2.79	64.27	2.65
(Child who lives with grandparents)	(4.28)	(3.80)	(89.56)	(10.77)
Household Migration status	1.79	−0.90	19.33	1.49
(Child who lives in nuclear family)	(1.99)	(1.64)	(48.11)	(7.00)
Household Migration status	2.52	0.42	13.07	−0.07
(North China)	(2.04)	(1.62)	(64.06)	(1.74)
Household Migration status	1.87	0.00	24.85	*N.A.*
(South China)	(1.76)	(1.55)	(45.25)	*N.A.*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$, [9] the regression results are not available due to missing values for the variables immunization shots and childcare. In some subsamples, the effective sample sizes for those two variables are too small to produce reliable regression results, where the effective sample contains the individuals that have more than one observation in the data.

Table 4.20b: Fixed effects model results of the effects of the household migration status on children's health outcome and care on subsamples: IV approach

	Calorie	Protein	Calorie/RDA	Protein/RDA
Household Migration status	3032.80	71.07	1.66	1.09
(Low income household)	(2878.64)	(70.86)	(1.66)	(1.22)
Household Migration status	−346.98	−18.25	−0.14	−0.25
(High income household)	(3812.96)	(126.21)	(2.66)	(2.66)
Household Migration status	1305.09	27.13	0.66	0.39
(Parents with low education level)	(931.02)	(25.21)	(0.56)	(0.47)
Household Migration status	1722.03*	48.37*	0.94·	0.77·
(Child above age 5)	(867.25)	(24.18)	(0.48)	(0.40)
Household Migration status	187.62	31.20	0.03	0.45
(Child who lives with grandparents)	(1696.92)	(55.35)	(1.12)	(1.05)
Household Migration status	1957.27	31.99	0.99	0.34
(Child who lives in nuclear family)	(1513.14)	(37.61)	(0.89)	(0.72)
Household Migration status	1653.92	29.98	0.78	0.35
(North China)	(1402.55)	(35.05)	(0.88)	(0.72)
Household Migration status	767.37	18.74	0.40	0.18
(South China)	(917.81)	(30.35)	(0.56)	(0.58)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Table 4.21a: Fixed effects model results of the effects of the father's migration status on children's health outcome and care on subsamples: IV approach

	WAZ	HAZ	Immunization shots	Childcare by non-family member
Father Migration status	2.49	1.43	76.41	−2.24
(Low income household)	(2.05)	(1.90)	(93.55)	(6.24)
Father Migration status	2.14	−1.77	*N.A.*	*N.A.*
(High income household)	(4.54)	(3.99)	*N.A.*[15]	*N.A.*
Father Migration status	2.07	2.28	68.83	*N.A.*
(Parents with low education level)	(1.34)	(1.51)	(69.92)	*N.A.*
Father Migration status	1.38	0.68	11.26	*N.A.*
(Child above age 5)	(0.91)	(0.92)	(25.00)	*N.A.*
Father Migration status	4.50	5.22	54.56	5.24
(Child who lives with grandparents)	(5.91)	(6.99)	(70.88)	(20.42)
Father Migration status	1.79	0.17	38.97	−0.68
(Child who lives in nuclear family)	(1.58)	(1.30)	(49.99)	(3.53)
Father Migration status	2.29	1.57	18.89	−3.07
(North China)	(1.92)	(1.84)	(75.66)	(7.08)
Father Migration status	2.42	0.85	37.96	*N.A.*
(South China)	(2.22)	(1.88)	(41.46)	*N.A.*

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$, [10] the regression results are not available due to missing values for the variables immunization shots and childcare. In some subsamples, the effective sample sizes for those two variables are too small to produce reliable regression results, where the effective sample contains the individuals that have more than one observation in the data.

Table 4.21b: Fixed effects model results of the effects of the father's migration status on children's health outcome and care on subsamples: IV approach

	Calorie	Protein	Calorie/RDA	Protein/RDA
Father Migration status	2029.88	45.86	1.10	0.64
(Low income household)	(1476.19)	(37.03)	(0.88)	(0.67)
Father Migration status	−1831.69	−89.18	−1.44	−2.19
(High income household)	(3122.73)	(127.10)	(2.30)	(2.97)
Father Migration status	1286.81	31.95	0.66	0.42
(Parents with low education level)	(859.56)	(24.21)	(0.52)	(0.44)
Father Migration status	1354.02$^{\cdot}$	36.92$^{\cdot}$	0.72$^{\cdot}$	0.54
(Child above age 5)	(749.43)	(20.53)	(0.41)	(0.33)
Father Migration status	−37.47	58.81	0.06	1.00
(Child who lives with grandparents)	(2240.88)	(87.28)	(1.49)	(1.65)
Father Migration status	1322.39	11.15	0.56	−0.18
(Child who lives in nuclear family)	(1009.21)	(26.99)	(0.60)	(0.58)
Father Migration status	1318.14	17.74	0.57	0.04
(North China)	(1284.92)	(32.55)	(0.83)	(0.70)
Father Migration status	886.57	21.65	0.41	0.02
(South China)	(1065.57)	(34.84)	(0.64)	(0.66)

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$.

Appendix C

Appendix to Chapter 2

A 2.1. Definition of variables

- Home ownership status: SIPP asks questions on the ownership status of the living quarters. Thus, homeowners are defined as those who own the living quarters, excluding outright owners. Renters are defined as those who are paying rent, excluding those who live in public housing.

- Labor market activities: SIPP contains data on monthly employment status[1] and labor income, which makes it possible to construct an individual's complete work history throughout the survey period. I use three variables to describe individuals' labor market outcomes:

  (1)Employment status (employed, unemployed and out of labor force): The individual is defined as unemployed as long he/she is unemployed for part of the interview period.[2] To avoid mis-coding non-participation as unemployment,

---

[1]In SIPP, the basic labor force information has been recoded into eight employment status recodes (ESR's). These ESRs are defined as follows:
ESR 1 –With job entire month, worked all weeks.
ESR 2 –With job entire month, missed 1 or more weeks, but not because of a layoff.
ESR 3 –With job entire month, missed 1 or more weeks because of a layoff.
ESR 4 –With job part of month, but not because of a layoff or looking for work.
ESR 5 –With job part of month, some time spent on layoff or looking for work.
ESR 6 –No job in month, spent entire month on layoff or looking for work.
ESR 7 –No job in month, spent part of month on layoff or looking for work.
ESR 8 –No job in month, no time spent on layoff or looking for work.

[2]This definition is consistent with that of the Bureau of Census. That is, unemployment consists of ESR=3 or ESR=5 or ESR=6 or ESR=7.

I dropped those who are out of the labor market from the sample.[3]

(2)Unemployment duration if unemployed: Unemployment duration measures how long people remained unemployed during each of the times (spells) they experienced unemployment. It's defined as an uninterrupted period of months in which an individual was unemployed. I use four months as the time period for unemployment spells since employment status is defined every four months. Using four months as the time unit might overestimate the unemployment duration compared with those using one month as the time unit since many spells shorter than four months have been recoded.

(3)Monthly wage if employed: Monthly wage is obtained by multiplying hourly wage rate by reported hours of labor supply per month whenever the worker is paid by the hour. For each month, respondents report hours of work per week and how many weeks worked. Monthly labor supply is calculated as hours per week(weeks worked/weeks in month)$\times$ 4.33. For those who are not paid hourly, their real wages are measured by their monthly earnings from the job.

A 2.2 Mortgage interest subsidy

The mortgage interest subsidy rate is calculated as a tax saving from an additional dollar of mortgage interest. I first calculate the state income tax liabilities owed by a representative sample of taxpayers in SIPP. [4] Then, I increase the mortgage interest by 1% for the taxpayer and recalculate the state taxes. Then, the mortgage interest subsidy rate is generated as the ratio of the additional tax (savings) to the additional 1% mortgage interest. The average mortgage interest subsidy of different

---

[3]I define those whose ESR=8 as non-participating. According to this definition, 32.3% of individuals are out of the labor market, 82% of whom were older than 65 and, therefore, retired.

[4]Representative taxpayers with high, media and low annual income has been tried and the main estimation results are kept as the same.

states in the period 1996-1999 [5] is presented in Table A1.

A 2.3 Regression Model for two-stage least squares analysis

The first stage estimates the following linear ownership equation:

$$H_{it} = \alpha_1 Z_{it} + \alpha_2 X1_{it} + \epsilon_{hit},$$

where H_{it} is a binary indicator where 1 indicates that the agent is a homeowner. Z_{it} is "mortgage interest subsidy." States with a higher "mortgage interest subsidy" indicate a more favorable tax code for homeowners. $X1_{it}$ are other controls that are related to ownership choices, including age, marriage status, whether the agent has a child(ren), whether the agent has a college education and year dummies.

In the second stage, I consider the following linear model:

$$U_{it} = \beta_1 \hat{H}_{it} + \beta_2 X2_{it} + \epsilon_{uit},$$

where U_{it} is a binary indicator with 1 indicating that the agent is unemployed. $\hat{H}_{it}$ is the predicted ownership indicator from the first stage. The instrumental variable is not correlated with the error term in the unemployment equation; that is, $E(\epsilon_{uit}|Z_i t) = 0$. Then, the causal effect is measured by β_1, which is the focus of the research. $X2_{it}$ are other controls that are correlated with unemployment.

A 2.4 Numerical solution of the dynamic model

The model is formulated as a dynamic program and solved numerically by backward recursion on the value function. As in Keane and Wolpin (1994), to cope with the "curse of dimensionality," given the value functions in period $t+1$, the value functions of t are calculated in three steps. In the first step, the expected

[5] I cannot identify all 50 states because in the SIPP 1996 panel, Maine and Vermont, as well as North Dakota and South Dakota share the same state code.

value functions at period $t+1$, the so-called "Emax," are simulated by Monte Carlo integration at a chosen subset of points in state space. In the second step, Emax at the remaining state points are calculated by linear regression interpolation. In the third step, Emax is used to calculate value function in period t. The three-step procedure is repeated backwards until value functions have been approximated for all $t = 0, \cdots, T$.

For illustrative purpose, let's consider the calculation of the expected future utility $Emax^{ul}$ associated with the condition that the unemployed agent receives a local job offer (the first part of the Bellman equation for an unemployed agent).

$$
\begin{aligned}
& Emax^{ul}_{t+1}(h_t, Z_{t+1}, X_{t+1}) \\
= & \int \max\left\{E_{S^-_{t+1}}[V^e_{t+1}(S^e_{t+1}), E_{S^-_{t+1}}[V^u_{t+1}(S^u_{t+1})\right\} dF(w) \\
= & \int\int \max\left\{V^e_{t+1}(h_t, w_{t+1}, Z_{t+1}, X_{t+1}, \eta_{ht+1}, \eta_{zt+1}, \eta_{wt+1}), V^u_{t+1}(h_t, Z_{t+1}, X_{t+1}, \eta_{ht+1}, \eta_{zt+1}, \eta_{wt+
\\
& dF_{\eta_{wt+1}}(w) dG(\eta_{ht+1}, \eta_{wt+1}, \eta_{zt+1}),
\end{aligned}
$$

where G is the joint distribution function of the three stochastic elements $\{\eta_{ht+1}, \eta_{wt+1}, \eta_{zt+1}\}$. Suppose that $\{h^l, Z^l, X^l\}$ represents the lth point in the chosen subset of state space, $\{\eta^r_{ht+1}, \eta^r_{wt+1}, \eta^r_{zt+1}\}$ represents the rth draws from the distribution of joint distribution and w^{rk} represents the kth draw from distribution $F_{\eta^r_{wt+1}}(w)$. Then, expected utility $Emax^{ul}$ at $\{h^l, Z^l, X^l\}$ can be approximated as below:

$$
\begin{aligned}
& Emax^{ul}_{t+1}(h^l, Z^l, X^l) \\
\approx & \frac{1}{R}\sum_{r=1}^{R} \max\left\{\frac{1}{K}\sum_{k=1}^{K} V^e_{t+1}(w^{rk}, h^l, Z^l, X^l, \eta^r_h, \eta^r_z, \eta^r_w) dF_{\eta^r_w}(w), V^u_{t+1}(h^l, Z^l, X^l, \eta^r_h, \eta^r_z, \eta^r_w)\right\} \\
& dG(\eta^r_h, \eta^r_w, \eta^r_z).
\end{aligned}
$$

In the same way, the other parts of the expected value functions can be calculated for all L points inside of the chosen state space. Then, the linear regression interpolation method is adopted to approximate the expected value functions at the remaining points of the state space. Given expected value functions, value functions are calculated as the maximum of lifetime utility of feasible choices.

A 2.5 Computation and Optimization Algorithms of Likelihood Function

Likelihood contribution L_i consists of different transition probabilities depending on the observable. For simplicity of expression, I categorize and express these transition probabilities into five employment transitions.

- From unemployment to local job employment: unemployed agent receives and accepts a local offer.

  $(E_t = 0, E_{t+1} = 1, m_{t+1} = 0)$

  Suppose that $\hat{h}$ and $\hat{w}$ are the observed home ownership status and employed wage. Then, the probability of observing $\{h_{t+1} = \hat{h}, E_{t+1} = 1, m_{t+1} = 0, w_{t+1} = \hat{w}\}$ given state $\{E_t = 0, h_t, w_t, X_t, Z_t, \bar{S}_t\}$ can be expressed as

  $$\begin{aligned} & P(h_{t+1} = \hat{h}, E_{t+1} = 1, m_{t+1} = 0, w_{t+1} = \hat{w} | E_t = 0, h_t, w_t, X_t, Z_t, \bar{S}_t) \\ = \ & \lambda_l \int \Big\{ f_{\eta_{wt+1}}(w = \hat{w}) \times Pr[d^{e\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})] \Big\} dG(S^{-}_{t+1}), \end{aligned}$$

  where $\lambda_l f_{\eta_{wt+1}}(w = \hat{w})$ stands for the probability of an agent receiving a local wage offer $\hat{w}$ when he/she is endowed with unobserved ability η_{wt+1}, and $Pr[d^{e\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})]$ stands for the probability that the agent accepts the wage offer and optimally chooses home ownership status $\hat{h}$.

- From unemployment to non-local job employment: unemployed agent receives and accepts a non-local offer.

  $(E_t = 0, E_{t+1} = 1, m_{t+1} = 1)$

  Suppose that $\hat{h}$ and $\hat{w}$ are the observed home ownership status and employed wage. Then, the probability of observing $\{h_{t+1} = \hat{h}, E_{t+1} = 1, m_{t+1} = 1, w_{t+1} =$

$\hat{w}\}$ given state $\{E_t = 0, h_t, w_t, X_t, Z_t, \bar{S}_t\}$ can be expressed as

$$\begin{aligned} & P(h_{t+1} = \hat{h}, E_{t+1} = 1, m_{t+1} = 1, w_{t+1} = \hat{w} | E_t = 0, h_t, w_t, X_t, Z_t, \bar{S}_t) \\ = \quad & \lambda_n \int \Big\{ f_{\eta_{wt+1}}(w = \hat{w}) \times Pr[d^{e\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})] \Big\} dG(S^{-}_{t+1}), \end{aligned}$$

where $\lambda_n f_{\eta_{wt+1}}(w = \hat{w})$ stands for the probability of an agent receiving a local wage offer $\hat{w}$ when he/she is endowed with unobserved ability η_{wt+1}; and, similarly, $Pr[d^{e\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})]$ stands for the probability that the agent accepts the wage offer and optimally chooses home ownership status $\hat{h}$.

- From unemployment to unemployment: unemployed agent has no offer or unemployed agent rejects the made offer.

  $(E_t = 0, E_{t+1} = 0, m_{t+1} = 0)$

  Suppose that $\hat{h}$ is the observed home ownership status. Then, the probability of observing $\{h_{t+1} = \hat{h}, E_{t+1} = 1, w_{t+1} = 0\}$ given state $\{E_t = 0, h_t, w_t, X_t, Z_t, \bar{S}_t\}$ can be expressed as

$$\begin{aligned} & P(h_{t+1} = \hat{h}, E_{t+1} = 0 | E_t = 0, h_t, w_t, X_t, Z_t, \bar{S}_t) \\ = \quad & (1 - \lambda_n - \lambda_l) + (\lambda_n + \lambda_l) \int \Big\{ Pr[d^{u\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})] \Big\} dG(S^{-}_{t+1}), \end{aligned}$$

  where $(1 - \lambda_n - \lambda_l)$ stands for the probability of an agent receiving no wage offer, and $(\lambda_n + \lambda_l) Pr[d^{u\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})]$ stands for the probability that the agent rejects any wage offer and optimally chooses home ownership status $\hat{h}$.

- From employment to employment: no layoff.

  $(E_t = 1, E_{t+1} = 1)$

  Suppose that $\hat{h}$ and $\hat{w}$ are the observed home ownership status and employed wage. Then, the probability of observing $\{h_{t+1} = \hat{h}, E_{t+1} = 1, w_{t+1} = \hat{w}\}$ given state $\{E_t = 1, h_t, w_t, X_t, Z_t, \bar{S}_t\}$ can be expressed as

$$
\begin{aligned}
& P(h_{t+1} = \hat{h}, E_{t+1} = 1, w_{t+1} = \hat{w} | E_t = 1, h_t, w_t, X_t, Z_t, \bar{S}_t) \\
= \quad & (1-\delta) \int \Big\{ Pr[d^{e\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})] \Big\} dG(\bar{S}_{t+1}),
\end{aligned}
$$

where $1-\delta$ stands for the probability that an agent stays employed, and $Pr[d^{e\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})]$ stands for the probability that the agent optimally chooses home ownership status $\hat{h}$.

- From employment to unemployment: layoff.

  $(E_t = 1, E_{t+1} = 0)$

  Suppose that $\hat{h}$ and $\hat{w}$ are the observed home ownership status and employed wage. Then, the probability of observing $\{h_{t+1} = \hat{h}, E_{t+1} = 0, w_{t+1} = \hat{w}\}$ given state $\{E_t = 1, h_t, w_t, X_t, Z_t, \bar{S}_t\}$ can be expressed as

$$
\begin{aligned}
& P(h_{t+1} = \hat{h}, E_{t+1} = 0, w_{t+1} = \hat{w} | E_t = 1, h_t, w_t, X_t, Z_t, \bar{S}_t) \\
= \quad & \delta \int \Big\{ Pr[d^{u\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})] \Big\} dG(\bar{S}_{t+1}),
\end{aligned}
$$

  where δ stands for the probability that an agent is laid off, and $Pr[d^{u\hat{h}}_{t+1} = \arg\max V_{t+1}(S_{t+1})]$ stands for the probability that the agent optimally chooses home ownership status $\hat{h}$.

The transition probabilities above contain high-dimensional integrals without a closed-form solution. I use the GHK simulator of Geweke (1991), Hajivassiliou (1991) and Keane (1994) to approximate the probability expression. Roughly speaking, this approach centers around the rewriting of the joint probability as a product of conditional probabilities and approximating this product by repeatedly computing marginal probabilities, drawing values of random variables consistent with conditions that must be satisfied, and updating conditional distributions. Unlike the crude frequency simulator, the GHK simulator is a continuous and differentiable function of the model's parameters. The smoothness permits the application of standard gradient optimization algorithms for maximum likelihood estimation and standard asymptotic theory for the proof of consistency and asymptotic normality of the estimation. The GHK simulator is also more efficient (in terms of the variance of the estimators of probabilities) than other simulators, such as the acceptance/rejection or Stern simulators.

The number of simulation draws R is set to equal 200, which satisfies the rule of thrum of setting R equal to an integer approximately equal to the square root of the sample size, and the estimates are likely to be insensitive to the choice of seed.

Gradient-based methods are widely used in log-likelihood optimization because many statistical models have well-behaved log-likelihood functions and gradients, and Hessian matrices are often required for statistical inference–e.g., in obtaining standard errors. The standard Newton-Raphson method, however, has well-known drawbacks. One is that computation of the Hessian matrix can be quite computationally intensive. Another problem is that the Hessian matrix may fail to be positive definite. To solve these two problems, I adopt the BHHH algorithm,[6] the idea of which is based on information matrix equality, replacing the Hessian by the negative of the sum over the outer products of the gradients of individual (independent) observations. This method yields significant computational benefits and guarantees a positive definite Hessian matrix (approximated). Finally, the inverse of the negative Hessian is the approximate variance-covariance matrix of the estimated parameters.

---

[6]See Berndt, Hall, and Hall (1974).

A 2.6 Algorithm for adaptive LASSO solution and K-fold cross validation criterion

The adaptive lasso can be solved by the same algorithm for solving the lasso. To get adaptive LASSO solutions, we need to optimize the simulated likelihood function subject to an inequality constraint $\sum_j |\hat{\theta}_j| < t$. This non-differentiable constraint can be converted into a set of linear constraints for all possible combination of signs of each estimated parameter, which can be easily incorporated into the BHHH algorithm for likelihood function. The number of the constraints is 2^J, with J as the total number of parameters, which is not tractable when J is large. To avoid this major drawback, I follow the method proposed by Tibshirani (1996). The basic idea behind this algorithm is to solve the likelihood function with just one constraint first and then check whether the solution satisfies $\sum_j |\hat{\theta}_j| < t$. If so, the computation is complete; if not, the violated constraint is added to the previous optimization problem and the process is continued until $\sum_j |\hat{\theta}_j| < t$. At termination, this algorithm must reach a solution that satisfies the constraints. This algorithm converges in approximately $0.5J$ to $0.75J$ iterations.

The value of turning parameter t plays an important role in controlling the regularization incorporated by the inequality constraint. If t is chosen larger than $\sum_j |\hat{\theta}_j|$, then the LASSO estimates are the θ_j. However, if t is set to be zero, then all parameters have to be zero. I choose the optimal turning parameter to minimize the prediction error for the purpose of prediction accuracy. The prediction error is estimated by the two dimensional cross-validation method. The basic steps of cross-validation include: (1) split the data into K roughly equal-sized sub-samples (folds); (2)for the k-th part, fit the model to the other $K-1$ parts of the data. That is, for each value combination of $(\hat{t}, \hat{\gamma})$,[7] calculate the prediction error of the fitted model

[7] The value of the turning parameter and adaptive parameter are discretized at 5 values.

when predicting the k-th part of the data. Denote by $\hat{f}^{-k}_{\hat{t},\hat{\gamma}}(x)$ the fitted function with $(\hat{t}, \hat{\gamma})$, computed with the k-th part of the data removed. Then, the estimate of prediction error is

$$e(\hat{f}_{\hat{t},\hat{\gamma}}) = \sum_{\text{i in k-th fold}} (y_i - \hat{f}^{-k}_{\hat{t},\hat{\gamma}}(x_i))^2;$$

(3)for each value of $(\hat{t}, \hat{\gamma})$, compute the average error over all folds; and (4) choose $(\hat{t}, \hat{\gamma})$ that minimize the cross-validation. Typical choices of K are five or ten, as mentioned in Hastie, Tibshirani, and Friedman (2001).

Table A1: NBER Mortgage Interest Subsidy Rate by US state in %, 1996-1999

U.S. State	*Mortgage Subsidy*	*Std. Dev.*	*Min.*	*Max.*	*No.of Obs.*
Alabama	3.94	0.31	3.65	4.28	1795
Alaska	0.00	0.00	0.00	0.00	366
Arizona	3.22	0.11	3.10	3.38	2892
Arkansas	6.69	0.00	6.69	6.69	974
California	6.62	0.92	6.00	8.00	20123
Colorado	4.95	0.10	4.63	5.00	2975
Connecticut	4.50	0.00	4.50	4.50	2462
Delaware	6.57	0.45	5.46	6.97	664
DC	9.30	0.00	9.30	9.30	35
Florida	0.00	0.00	0.00	0.00	7323
Georgia	5.66	0.00	5.66	5.66	4556
Hawaii	8.99	0.67	7.81	9.37	108
Idaho	8.20	0.01	8.10	8.20	1086
Illinois	3.00	0.00	3.00	3.00	8316
Indiana	3.40	0.00	3.40	3.40	5833
Iowa	6.84	0.40	5.89	7.47	2082
Kansas	6.25	0.00	6.25	6.25	2007
Kentucky	6.00	0.00	6.00	6.00	3219
Louisiana	2.85	0.01	2.85	2.90	1866
Maryland	4.95	0.07	4.85	5.00	2033
Massachusetts	5.95	0.01	5.85	5.95	3946
Michigan	4.40	0.02	4.20	4.40	6309
Minnesota	7.83	0.32	7.05	8.00	3772
Mississippi	4.96	0.07	4.84	5.00	1118
Missouri	4.90	0.34	4.39	5.15	3739
Montana	7.85	0.48	6.66	8.63	696
Nebraska	10.81	0.19	10.40	10.97	826
Nevada	0.00	0.00	0.00	0.00	639
New Hampshire	0.00	0.00	0.00	0.00	885
New Jersey	5.52	0.00	5.52	5.53	5566
New Mexico	7.10	0.00	7.10	7.10	724
New York	8.36	0.12	8.17	8.44	9608
North Carolina	7.00	0.00	7.00	7.00	3950
Ohio	6.15	0.13	5.95	6.32	7253
Oklahoma	6.49	0.09	6.33	6.54	2116
Oregon	9.00	0.00	9.00	9.00	1684
Pennsylvania	2.80	0.00	2.80	2.80	9027
Rhode Island	5.75	1.70	3.75	7.41	635
South Carolina	7.00	0.00	7.00	7.00	1736
Tennessee	0.00	0.00	0.00	0.00	3067
Texas	0.00	0.00	0.00	0.00	11018
Utah	6.09	0.00	6.09	6.09	1667
Virginia	5.65	0.01	5.65	5.75	3729
Washington	0.00	0.00	0.00	0.00	3612
West Virginia	6.50	0.00	6.50	6.50	1519
Wisconsin	6.86	0.09	6.55	6.93	4242
Maine, Vermont	7.27	0.97	6.50	8.50	1157
North Dakota, South Dakota	1.89	1.87	0.00	3.77	1099

Note: Mortgage Interest Subsidy Rate is calculated by TAXSIM provided by NBER.

Table A2: Full list of 20 possible transitions

Transition from unemployment	3*4=12 possibilities
to local jobs, rent to rent	$(E_t = 0, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 0)$
to local jobs, rent to own	$(E_t = 0, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 0)$
to local jobs, own to rent	$(E_t = 0, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 0)$
to local jobs, own to own	$(E_t = 0, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 0)$
to non-local jobs, rent to rent	$(E_t = 0, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 1)$
to non-local jobs, rent to own	$(E_t = 0, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 1)$
to non-local jobs, own to rent	$(E_t = 0, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 1)$
to non-local jobs, own to own	$(E_t = 0, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 1)$
to unemployment, rent to rent	$(E_t = 0, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 0)$
to unemployment, rent to own	$(E_t = 0, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 0)$
to unemployment, own to rent	$(E_t = 0, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 0)$
to unemployment, own to own	$(E_t = 0, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 0)$
Transition from employment	2*4=8 possibilities
to unemployment, rent to rent	$(E_t = 1, h_t = 0 \mid E_{t+1} = 0, h_{t+1} = 0, m_{t+1} = 0)$
to unemployment, rent to own	$(E_t = 1, h_t = 0 \mid E_{t+1} = 0, h_{t+1} = 1, m_{t+1} = 0)$
to unemployment, own to rent	$(E_t = 1, h_t = 1 \mid E_{t+1} = 0, h_{t+1} = 0, m_{t+1} = 0)$
to unemployment, own to own	$(E_t = 1, h_t = 1 \mid E_{t+1} = 0, h_{t+1} = 1, m_{t+1} = 0)$
to employment, rent to rent	$(E_t = 1, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 0)$
to employment, rent to own	$(E_t = 1, h_t = 0 \mid E_{t+1} = 1, h_{t+1} = 01, m_{t+1} = 0)$
to employment, own to rent	$(E_t = 1, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 0, m_{t+1} = 0)$
to employment, own to own	$(E_t = 1, h_t = 1 \mid E_{t+1} = 1, h_{t+1} = 1, m_{t+1} = 0)$

Bibliography

ABBRING, J. H. (2010): "Identification of Dynamic Discrete Choice Models," *Annual Review of Economics*, 2(1), 367–394.

ANGRIST, J. D. (2001): "Estimation of limited dependent variable models with dummy endogenous regressors," *Journal of Business & Economic Statistics*, 19(1).

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 91(434), 444–455.

BAO, S., Ö. B. BODVARSSON, J. W. HOU, AND Y. ZHAO (2009): "Migration in China from 1985 to 2000," *Chinese Economy*, 42(4), 7–28.

BATTU, H., A. MA, AND E. PHIMISTER (2008): "Housing tenure, job mobility and unemployment in the UK," *Economic Journal*, 118(527), 311–328.

BERNDT, E. K., B. H. HALL, AND R. E. HALL (1974): "Estimation and inference in nonlinear structural models," *Annals of Economic and Social Measurement*, 3(4), 103–116.

BLANCHFLOWER, D. G., AND A. J. OSWALD (2013): "Does High Home-Ownership Impair the Labor Market?," *NBER Working paper*.

BLAU, D. (2011): "Pensions, household saving, and welfare: A dynamic analysis," *IZA Discussion Paper*.

Blundell, R., and T. MaCurdy (1999): "Labor supply: A review of alternative approaches," *Handbook of labor economics*, 3, 1559–1695.

Bottazzi, R. (2004): "Labour market participation and mortgage-related borrowing constraints," .

Burdett, K., and D. T. Mortensen (1980): "Search, layoffs, and labor market equilibrium," *The Journal of Political Economy*, pp. 652–672.

Cai, F., A. Park, and Y. Zhao (2004): "The Chinese Labor Market in the Reform Era," *This page intentionally left blank*, p. 167.

Campbell, J. Y., and J. F. Cocco (2007): "How do house prices affect consumption? Evidence from micro data," *Journal of Monetary Economics*, 54(3), 591–621.

Carroll, C. D. (1997): "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *The Quarterly Journal of Economics*, 112(1), 1–55.

Carroll, C. D., and M. S. Kimball (2001): "Liquidity constraints and precautionary saving," Discussion paper, National Bureau of Economic Research.

Chen, S. H. (2008): "Estimating the variance of wages in the presence of selection and unobserved heterogeneity," *The Review of Economics and Statistics*, 90(2), 275–289.

Chen, X., Q. Huang, S. Rozelle, Y. Shi, and L. Zhang (2009): "Effect of migration on children's educational performance in rural China," *Comparative Economic Studies*, 51(3), 323–343.

Chetty, R. (2006): "A new method of estimating risk aversion," *The American Economic Review*, pp. 1821–1834.

Chetty, R., and A. Szeidl (2010): "Consumption commitments: a foundation for reference-dependent preferences and habit formation," *Working Paper*.

Coulson, N., and L. Fisher (2002): "Tenure choice and labour market outcomes," *Housing Studies*, 17(1), 35–49.

——— (2009): "Housing tenure and labor market impacts: The search goes on," *Journal of Urban Economics*, 65(3), 252–264.

Cunha, F., and J. Heckman (2007): "The technology of skill formation," Discussion paper, National Bureau of Economic Research.

De Brauw, A. (2008): *Policy Research Working Paper Migrant Labor Markets and the Welfare of Rural Households in the Developing World: Evidence from China*, vol. 4585. World Bank Publications.

De Brauw, A., and R. Mu (2011): "Migration and the overweight and underweight status of children in rural China," *Food Policy*, 36(1), 88–100.

DiPasquale, D., and E. L. Glaeser (1999): "Incentives and social capital: are homeowners better citizens?," *Journal of urban Economics*, 45(2), 354–384.

Du, S., T. A. Mroz, F. Zhai, and B. M. Popkin (2004): "Rapid income growth adversely affects diet quality in Chinaparticularly for the poor!," *Social science & medicine*, 59(7), 1505–1515.

Eckstein, Z., and G. Van den Berg (2007): "Empirical labor search: A survey," *Journal of Econometrics*, 136(2), 531–564.

Eckstein, Z., and K. I. Wolpin (1999): "Why youths drop out of high school: The impact of preferences, opportunities, and abilities," *Econometrica*, 67(6), 1295–1339.

Feenberg, D., and E. Coutts (1993): "An introduction to the TAXSIM model," *Journal of Policy Analysis and Management*, 12(1), 189–194.

Ferreira, F., J. Gyourko, and J. Tracy (2010): "Housing busts and household mobility," *Journal of Urban Economics*, 68(1), 34–45.

Flatau, P., M. Forbes, and P. Hendershott (2003): "Homeownership and unemployment: the roles of leverage and public housing," *NBER Working paper*.

Flavin, M., and S. Nakagawa (2008): "A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence," *The American economic review*, pp. 474–495.

Flinn, C., and J. Heckman (1982): "New methods for analyzing structural models of labor force dynamics," *Journal of Econometrics*, 18(1), 115–168.

Fortin, N. M. (1995): "Allocation inflexibilities, female labor supply, and housing assets accumulation: are women working to pay the mortgage?," *Journal of Labor Economics*, 13(3), 524–57.

French, E., and C. Taber (2011): "Identification of models of the labor market," *Handbook of Labor Economics*, 4, 537–617.

Gemici, A. (2007): "Family migration and labor market outcomes," *Working paper*.

Geweke, J. (1991): "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities," *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pp. 571–578.

Giles, J. (2006): "Is life more risky in the open? Household risk-coping and the opening of China's labor markets," *Journal of Development Economics*, 81(1), 25–60.

GLAESER, E. L., AND J. M. SHAPIRO (2003): "The Benefits of the Home Mortgage Interest Deduction," *Tax Policy and the Economy*, pp. 37–82.

GREEN, R., AND P. HENDERSHOTT (2001): "Home-ownership and unemployment in the U.S.," *Urban Studies*, 38(9), 1509.

HAJIVASSILIOU, V. (1991): "Simulation Estimation Methods for Limited Dependent Variable Models," *Cowles Foundation discussion paper*.

HASTIE, T., R. TIBSHIRANI, AND J. J. H. FRIEDMAN (2001): *The elements of statistical learning*, vol. 1. Springer New York.

HEAD, A., AND H. LLOYD-ELLIS (2012): "Housing liquidity, mobility, and the labour market," *The Review of Economic Studies*, 79(4), 1559–1589.

HECKMAN, J. (1974): "Shadow prices, market wages, and labor supply," *Econometrica: journal of the econometric society*, pp. 679–694.

HECKMAN, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, pp. 153–161.

——— (2008): "Role of income and family influence on child outcomes," *Annals of the New York Academy of Sciences*, 1136(1), 307–323.

HECKMAN, J. J., S. H. MOON, R. PINTO, P. A. SAVELYEV, AND A. YAVITZ (2010): "The rate of return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 94(1), 114–128.

HECKMAN, J. J., AND S. URZUA (2010): "Comparing IV with structural models: What simple IV can and cannot identify," *Journal of Econometrics*, 156(1), 27–37.

HECKMAN, J. J., AND E. J. VYTLACIL (2007): "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," *Handbook of econometrics*, 6, 4779–4874.

Hilber, C. A., and T. M. Turner (2010): "The mortgage interest deduction and its impact on homeownership decisions," *Review of Economics and Statistics*, (0).

Karahan, F., and S. Rhee (2012): "Geographical relocation and unemployment duration," *Working paper*.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 77(1), 135–175.

Keane, M., and P. K. W. Todd (2010): "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," *Handbook of Labor Economics*, 4.

Keane, M. P. (1994): "A computationally practical simulation estimator for panel data," *Econometrica: Journal of the Econometric Society*, pp. 95–116.

——— (2010): "Structural vs. atheoretic approaches to econometrics," *Journal of Econometrics*, 156(1), 3–20.

Keane, M. P., and K. I. Wolpin (1994): "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence," *The Review of Economics and Statistics*, pp. 648–672.

——— (1997): "The career decisions of young men," *Journal of political Economy*, 105(3), 473–522.

Killingsworth, M. R., and J. J. Heckman (1986): "Female labor supply: A survey," *Handbook of labor economics*, 1(1), 103–204.

Liang, Z., and Z. Ma (2004): "China's floating population: new evidence from the 2000 census," *Population and development review*, 30(3), 467–488.

Liu, H., H. Fang, and Z. Zhao (2013): "Urban–rural disparities of child health and nutritional status in China from 1989 to 2006," *Economics & Human Biology*, 11(3), 294–309.

Magnac, T., and D. Thesmar (2002): "Identifying dynamic discrete decision processes," *Econometrica*, 70(2), 801–816.

Mallee, H. (1995): "China's household registration system under reform," *Development and Change*, 26(1), 1–29.

Modestino, A. S., and J. Dennett (2012): "Are American homeowners locked into their houses? The impact of housing market conditions on state-to-state migration," *Regional Science and Urban Economics*.

Moffitt, R. (2003): "Causal analysis in population research: An economist's perspective," *Population and Development Review*, 29(3), 448–458.

Morissette, R., and F. Hou (2008): "Does the labour supply of wives respond to husbands' wages? Canadian evidence from micro data and grouped data," *Canadian Journal of Economics/Revue canadienne d'économique*, 41(4), 1185–1210.

Mu, R., and D. van de Walle (2011): "Left behind to farm? Women's labor re-allocation in rural China," *Labour Economics*, 18, S83–S97.

Munch, J. R., M. Rosholm, and M. Svarer (2006): "Are Homeowners Really More Unemployed?," *The Economic Journal*, pp. 991–1013.

Osberg, L., J. Shao, and K. Xu (2009): "The growth of poor children in China 1991–2000: why food subsidies may matter," *Health economics*, 18(S1), S89–S108.

Oswald, A. (1996): "A conjecture on the explanation for high unemployment in the industrialised nations," *Warwick Economic Research Papers.*

——— (1997): "The missing piece of the unemployment puzzle," *Working paper.*

Popkin, B. M., S. Du, F. Zhai, and B. Zhang (2010): "Cohort Profile: The China Health and Nutrition Surveymonitoring and understanding socio-economic and health change in China, 1989–2011," *International journal of epidemiology*, 39(6), 1435–1440.

Postlewaite, A., L. Samuelson, and D. Silverman (2004): "Consumption, Commitmants and Preferences for Risk," Discussion paper, National Bureau of Economic Research.

Rendon, S. (2006): "Job Search and asset accumulation under borrowing constraints," *International Economic Review*, 47(1), 233–263.

Rosen, H. S. (1979): "Housing decisions and the US income tax: An econometric analysis," *Journal of Public Economics*, 11(1), 1–23.

Rozelle, S., L. Guo, M. Shen, A. Hughart, and J. Giles (1999): "Leaving China's farms: survey results of new paths and remaining hurdles to rural migration," *The China Quarterly*, 158, 367–393.

Rust, J. (1994): "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4(4).

Schulhofer-Wohl, S. (2011): "Negative equity does not reduce homeowners' mobility," *NBER Working Paper.*

Semykina, A., and J. M. Wooldridge (2010): "Estimating panel data models in the presence of endogeneity and selection," *Journal of Econometrics*, 157(2), 375–380.

Shen, T., J.-P. Habicht, and Y. Chang (1996): "Effect of economic reforms on child growth in urban and rural areas of China," *New England Journal of Medicine*, 335(6), 400–406.

Shore, S. H., and T. Sinai (2010): "Commitment, Risk, and Consumption: Do Birds of a Feather Have Bigger Nests?," *The Review of Economics and Statistics*, 92(2), 408–424.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business & Economic Statistics*, 20(4).

Svedberg, P. (2006): "Declining child malnutrition: a reassessment," *International Journal of Epidemiology*, 35(5), 1336–1346.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, pp. 267–288.

Valletta, R. G. (2013): "House lock and structural unemployment," *Labour Economics*.

Van Leuvensteijn, M., and P. Koning (2004): "The effect of home-ownership on labor mobility in the Netherlands," *Journal of Urban Economics*, 55(3), 580–596.

Winkler, H. (2010): "The Effect of Homeownership on Geographic Mobility and Labor Market Outcomes," *Working paper*.

Wooldridge, J. M. (2005): "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of applied econometrics*, 20(1), 39–54.

——— (2010): *Econometric analysis of cross section and panel data*. MIT press.

Yao, R., and H. Zhang (2005): "Optimal consumption and portfolio choices with risky housing and borrowing constraints," *Review of Financial Studies*, 18(1), 197.

Yoshikawa, H., and F. Ohtaka (1989): "An analysis of female labor supply, housing demand and the saving rate in Japan," *European Economic Review*, 33(5), 997–1023.

Zhang, S. (2012): "Migration and Childrens Health: Evidence From Rural China,"
.

Zhao, Y. (1999): "Leaving the countryside: rural-to-urban migration decisions in China," *American Economic Review*, pp. 281–286.

Zou, H. (2006): "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, 101(476), 1418–1429.

Curriculum Vitae

Xi Yang was born in Xi'an, China on July 23, 1983. She received her B.S. in Mathematics and B.A. in Economics at Wuhan University, Wuhan, China in 2005. Xi entered the China Center for Economic Research at Peking University and earned her M.A.in 2008. Xi started her study in the Ph.D. program of economics at Johns Hopkins University from the fall of 2008. She will start as a postdoctoral scholar at the Lusk center for real estate at University of Southern California from August 2014.