5-2013

Essays on Agglomeration, Homeownership, and Labor Market Outcomes

Jie Dai
Syracuse University

Abstract

This dissertation consists of two essays on the labor market impact of agglomeration economies and that of homeownership. The empirical analyses use different U.S. data and econometric techniques to examine the impact on different labor market outcomes such as volatility of hours worked and men's employment.

The first essay is motivated by the labor market pooling model from Krugman (1991). The paper adds to a small but important literature that provides evidence on the microeconomic foundation of agglomeration economies. Using various years of data from the American Community Survey and the County Business Patterns survey, I show that the agglomeration of economic activities reduces the volatility of hours worked. Drawing on Krugman's model, I argue that this implies that labor pooling contributes to agglomeration economies, and helps to explain why cities are productive places.

In the second essay, I consider the extent to which homeownership affects men's labor supply. Research on the labor-supply consequences of homeownership is complicated by the endogeneity of housing tenure status. To address this endogeneity problem, I use a set of family size instruments (the presence of the third and additional children in a single family household) to estimate the effect of homeownership in a bivariate probit model. Based on a sample of married white male household heads from the American Community Survey, the IV result suggests that men who own their homes are 1.2% more likely to be employed relative to those who rent. I also show that the relationship between homeownership and family size is highly nonlinear and nonmonotonic. The first two children have positive influence on both homeownership and men's employment. The third

and additional children are negatively associated with homeownership but have no significant incremental effects on men's labor supply.

Essays on Agglomeration, Homeownership, and Labor Market Outcomes

By

Jie Dai

B.S., Zhejiang University, 2002
M.A., Syracuse University, 2008

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Graduate School of Syracuse University

Syracuse University
May 2013

Acknowledgements

My deepest gratitude goes to all of whom helped me complete this dissertation and contribute to my intellectual progress at Syracuse University.

In particular, I would like to thank my primary advisor, Stuart Rosenthal for his mentoring on my graduate study. He inspired me on these topics and patiently guided me through my dissertation. I would also like to thank Professor Jeffrey Kubik and Professor Chihwa Kao for their encouragement, support and insightful guidance in each step of my progress. My appreciation goes to Professor Stuart Brown and Professor Mary Lovely for their encouragement and support. I would also like to acknowledge the contribution of Professor Yildiray Yildirim to my dissertation defense.

I also extend my thankfulness to Professors Badi Baltagi, Donald Dutkowsky, Susan Gensemer, William Horrace, Jerry Kelly, Thomas Kniesner, Derek Laing, Jan Ondrich, and John Yinger, for helping build my knowledge and techniques for theoretic and applied economic research. I also benefit a lot from many constructive conversations with my fellow classmates and their great help. Moreover, I would like to acknowledge assistance from Sue Lewis, Laura Sauta in the Economics Department of Syracuse University.

Last, but most important, my lifelong appreciation goes to my family including my parents, my parents in law, and especially my wife Yu Shi. It is their love and belief in me that make this dissertation possible.

Table of Contents

List of Tables and Figures

Tables

Essay I: Agglomeration and Volatility of Hours Worked

Essay II: Homeownership and Men's Labor Supply: An IV Approach

Figures

Essay I: Agglomeration and Volatility of Hours Worked

Essay II: Homeownership and Men's Labor Supply: An IV Approach

Essay I: Agglomeration and Volatility of Hours Worked

1. Introduction

Labor market pooling is one of the many ways in which spatial concentrations of economic activities, also known as agglomeration economies, can increase productivity (Rosenthal and Strange, 2004; Combes, et al., 2007). Spatially concentrated pools of skilled labor are thought to provide spillover benefits that reduce job search cost and improve job match quality (Helsely and Strange, 1990). Another spillover from labor pooling is that it helps employers to adapt to firm-specific shocks, and reduces the volatility of wages and hours worked. The latter type of spillover was first mentioned in Marshall (1920)[1], and more carefully articulated by Krugman (1991, Ch. 2 and App. B). Krugman argued that in the presence of firm-specific shocks, employers operating close to pools of skilled labor had greater ability to vary the size of their workforce (through hiring and firing) while keeping hours worked stable. Employers operating in more isolated locations, as Krugman argued, were more reliant on overtime work of employed workers as a mechanism to adjust the level of labor input, thus their hours worked tend to have greater volatility.

Despite the theoretical predictions, the connection between agglomeration and the volatility of hours worked has received little attention in empirical literature. Only a few studies had offered evidence on the Krugman (1991) type of labor pooling spillover by examining the locations of companies of specific industries. Gerlach et al. (2009) found that US firms engaging in risky R&D activities tend to locate together to gain access to large pools of skilled labor. Similar patterns were found by Overman and Puga (2010) using UK data. They showed that the

[1] According to Marshall (1920), "A localized industry gains great advantage from the fact that it offers constant market for skill.…(whereas) the owner of an isolated factory is often put to great shifts for want of special skilled labor…"

industrial sectors whose establishments experience more idiosyncratic shocks tend to be more spatially concentrated. However, no empirical studies have ever considered the relationship between agglomeration and volatility of hours worked.

This paper examines the extent to which agglomeration reduces the volatility of hours worked through labor market pooling. The paper makes three contributions. First, it establishes that there is a systematic negative relationship between agglomeration and the volatility of hours worked. Second, it shows that this reduction in volatility of hours worked is larger among the skilled professional workers, relative to the less skilled nonprofessional workers. The third contribution of this paper is to provide evidence that helps to identify labor market pooling as a potential source of agglomeration economies. The empirical evidence presented in this paper is consistent with the predictions from the labor pooling model developed by Krugman (1991). To the best of my knowledge, this paper is the first study to provide evidence on hours worked in support of the model.

One of the many challenges in this study is to obtain data sufficient to test the spillover effect of labor pooling. The ideal data would be an employer-employee matched micro dataset. However, this type of data is usually confidential and I do not have access to it at this time. Instead, I use publicly available data from two sources, one is the American Community Survey (ACS) and the other is the County Business Patters (CBP) survey. The ACS data provides rich information on worker's socio-demographic attributes and labor market outcomes, while the CBP data provides information on the employment sizes and locations of business establishments. Data samples from these two sources are merged at the Metropolitan Statistical Area (MSA) level, and cover sample years from 2005 to 2009, and separately for year 2000[1].

Using the datasets from ACS and CBP, I use two different approaches to examine agglomeration's impact on volatility of hours worked. The first approach focuses on how inter-temporal change of average hours worked varies with degrees of labor pooling. Krugman (1991) suggested that in pooled settings, companies could easily hire and fire workers with a large pool of skilled workers nearby, therefore hours worked were more stable across time. In contrast, in isolated settings with little pooling, companies are more reliant on overtime work as a mechanism to vary labor input, which caused hours worked to be more variable across time. Moreover, in locations where the degrees of labor pooling are somewhere in between pooled settings and isolated settings, companies have to adjust their labor input through both mechanisms of hiring & firing and overtime work. In this case, wages and hours worked will vary across time at intermediate levels. By comparing across the three types of settings, I hypothesize that there is a negative relationship between the inter-temporal change of average hours worked and the degrees of labor pooling.

To test this hypothesized relationship described above, I start with a simple cross sectional analysis by regressing the standard deviation of MSA's average hours worked on a set of MSA agglomeration variables. The dependent variable measures how much the average hours worked of all workers in an MSA varies across sample years. The key MSA agglomeration variable is the average number of establishments in the MSA over sample years, which is a proxy for degrees of labor pooling. Following the cross sectional analysis, I carry out a panel

[1] For sample year 2000, data on workers demographics and labor market outcomes are from Census 2000 5% data, also available in the IPUMS database.

data analysis based on an MSA-level panel. I regress the first-differenced MSA average hours worked on a set of time-variant MSA agglomeration variables, including the number of establishments in the MSA. MSA fixed effects are included in the regressions to control for unobserved time-invariant MSA characteristics. Estimates from both the cross sectional analysis and panel analysis show a significant and negative relationship between inter-temporal change of hours worked and degrees of labor pooling.

My second approach is to examine how agglomeration reduces the cross sectional variance of individual's hours worked through labor pooling. As suggested by Krugman (1991), companies in more pooled settings have greater ability to vary the size of their workforce (through hiring and firing), and workers employed by these companies work for a competitive market wage rate and a standard number of hours worked. The implication is that the greater degree of labor pooling in an MSA, the smaller cross sectional variance of hours worked of all employed workers in the MSA. In contrast, because companies in more isolated settings are more reliant on overtime work as a mechanism to vary labor input, workers employed by different companies may work different numbers of hours. This suggests that MSAs with limited labor pooling should display relative larger cross sectional variance of hours worked of all employed workers in the MSA. The comparison between pooled settings and isolated settings shows that the cross sectional variance of hours worked tends to be smaller in MSAs with greater degrees of labor pooling, relative to MSAs with little labor pooling.

I use several different specifications and data to test the following hypothesis: the cross sectional variance of hours worked is negatively associated with degrees of labor pooling. In the main specification, I first regress individuals' hours worked on a set of socio--demographic controls, and generate the residuals as conditional hours worked. Next I calculate the cross sectional variance of these residuals in each MSA. This MSA-specific cross sectional variance of hours worked is then regressed on a set of MSA agglomeration variables, including a proxy for degrees of labor pooling. Based on the data from sample year 2000, the results show that the cross sectional variance of hours worked in an MSA has a negative relationship with the degrees of labor pooling in the MSA. In addition to the main specification, I experiment with alternative specification (using WPUMA based cross sectional variance) and data from different sample year (ACS 2008). The results are consistent with findings from my main specification.

More importantly, in this paper I also examine the differential impact of labor pooling on skilled professional workers versus less-skilled nonprofessional workers. Krugman (1991) argued that the spillover effects of labor pooling arise primarily from localized industries that rely on workers with industry-specific skills. This is both because companies require more skilled workers when they experience a positive market demand shock, and because it is expensive for companies to train new skilled workers or to attract skilled workers from other locations. The less skilled workers, however, are relative easier to train or replace. As a result, pooling of less-skilled workers provides little benefit to both companies and less-skilled workers. Drawing on the above comparison between skilled and less-skilled workers, I expect that the spillover effects from labor pooling have greater impact on skilled workers, relative to the less-skilled workers.

Throughout the paper, all of the regressions are run twice using data stratified by workers skill types, one for skilled professional workers and the other for less-skilled nonprofessional workers. The estimates show that the reductions in volatility of hours worked through labor pooling are consistently more significant and larger in magnitude for skilled workers, relative to the less-skilled workers. This result is congruent with the assumption in Krugman's (1991) model, that the spillover effect of labor pooling is primarily based only on skilled workers, but not on the less-skilled workers.

This paper is related to a small but important literature that provides evidence on the microeconomic foundations of agglomeration economies. There are many potential sources of agglomeration economies[1], and identifying the specific microfoundation of each source has been a challenging task for researchers. This is because these different sources all generate positive and similar productivity gains, making it difficult to trace out their specific mechanisms using productivity data[2]. Previous studies have resorted to alternative approaches and datasets, and have successfully identified some of the potential sources[3]. However, there is very limited evidence on how labor market pooling reduces the employment and unemployment risk associate with idiosyncratic firm-specific shocks (Gerlach et al., 2009; Overman and Puga, 2010). This paper adds to this literature and provides new evidence on this spillover from labor pooling by examining the reduction in volatility of hours worked in agglomeration economies.

The rest of the paper is organized in the following sequence. Section 2 introduces the data and key variables. Section 3 examines on how inter-temporal change of average hours worked varies with degrees of labor pooling, and reports estimation results based on the MSA-level panel data. Section 4 investigates on how cross sectional variance of hours worked varies with degrees of labor pooling, and reports estimation results based on cross sectional datasets. Section 5 concludes.

---

1 Such as intermediate input sharing, knowledge spillover, and labor market pooling, as suggested by Alfred Marshall (1920). In addition to these three, there are also many other potential sources proposed by previous studies, see Rosenthal and Strange (2004).

2 This is known as the "Marshallian equivalence", that all of the potential sources of agglomeration economies generate positive productivity gains, making it difficult to trace out specific mechanisms by measuring the quantity of the productivity gains. For more details, see survey by Duranton and Puga (2004).

3 See a survey of empirical studies on micro-foundations of agglomeration economies by Rosenthal and Strange (2004).

2. Data and Variables

In this paper, I use the IPUMS[1] 2000 Census and American Community Survey[2] (ACS) from 2005 to 2009 as the primary data for information on workers. The primary data source of business establishments for this study is the County Business Patterns Survey[3] (CBP), which is available to public through the US Census Bureau. I consider each MSA as an agglomeration economy in this study, and merge the individual-level records from the Census and ACS with the firms-level records from CBP at the MSA level, using their geographic codes. This approach allows me to proxy the degree of labor pooling by the total number of establishments in each MSA. It is worth noting that the data used in this paper is not employer-employee matched. An employer-employee matched data would be the ideal one for this study, however it is usually confidential and I do not have access to it at this time.

The ACS and Census data provide rich information on worker's socio-demographics and labor market outcomes, such as worker's *Usual Hours Worked Per Week*[4]. I restrict the data to include only male workers resides in MSAs, between age 25 and 55, worked for more than 50 weeks last year, and are employed in the private sectors. The full sample is stratified into two subsamples as skilled professional workers and nonprofessional workers with less skill. The professional subsample includes workers who have education attainments at the Master, Professional, or PhD degree level, and have their primary occupations in the Census professional

---

[1] The Integrated Public Use Microdata Series (IPUMS) website (www.ipums.org) provides a more user-friendly version of micro-data samples of United States census records for public use.
[2] The American Community Survey is conducted annually by the US Census Bureau.
[3] The County Business Patterns survey is conducted annually by the US Census Bureau, and public available on the Census Bureau website (www.census.gov).
[4] In the IPUMS this is measured using variable UHRSWORK, I include the workers with full length of UHRSWORK from 1 hour to maxim 99 hours.

occupation categories[1]. The nonprofessional subsample includes workers who have education attainments equal to or below high school graduate or GED, and have Census occupation types in all occupation categories other than the professional ones[2].

The CBP data is a good source of information on locations and employment sizes of all business establishments in the US. I restrict my sample to only large establishments that have more than 50 employees and locate in MSAs. Business establishments with less than 50 employees are mostly family and startup companies, whose employment decision in hiring and firing may not directly relate to market conditions and business shocks, and may be different from the standard competitive profit-maximizing business establishments. Exclusion of these small companies helps to reduce potential noise in the sample. One thing worth noting is that the CBP survey is county-coded, which cannot fully identify all coded MSAs in the IPUMS database. In several cases, a county straddles over the MSA boundary and extends to two adjacent MSAs[3]. To solve this problem, I allocate all workers and business establishments from a straddling county to the MSA where the county's majority population resides in.

Based on the combined data described above, I construct two separate datasets, one in panel setting and the other in cross sectional setting. The panel dataset follows 275 MSAs for five ACS sample years from 2005 to 2009, and it is used to test how inter-temporal change of

---

[1] Worker's occupation is classified according to the 1998 Standard Occupational Classification (SOC) system. This paper uses the first two digit code of the SOC code to categorize work's occupation. The professional occupations include the following categories that can be identified by the first 2 digit SOC code: Management, Business Operations Specialists, Financial Specialists, Computer and Mathematical Occupations, Architecture and Engineering Occupations, Life, Physical, and Social Science Occupations, Legal Occupations, Education, Training, and Library Occupations, Healthcare Practitioners and Technical Occupations.

[2] The nonprofessional group's first two digit SOC occupation categories includes: Food Preparation and Serving Occupations, Building and Grounds Cleaning and Maintenance, Sales Occupations, Office and Administrative Support Occupations, Construction Trades, Extraction Workers, Installation, Maintenance, and Repair Workers, Production Occupations, Transportation and Material Moving Occupations.

[3] There are eight counties that straddle over two MSAs: Middlesex County, MA; Norfolk County, MA; Hillsborough County, NH; New Haven County, CT; Fairfield County, CT; Plymouth County, MA; Worcester County, MA; Bristol County, MA.

MSA's average hours worked varies with degrees of labor pooling. The cross sectional dataset is a combination a sample of workers drawn from 2000 Census 5% file and a sample of business establishments from 2000 CBP survey, matched at MSA level. This cross sectional dataset is used to test how MSA's variance of individual hours worked varies with degrees of labor market pooling.

The summary statistics of all agglomeration variables are presented in Table 1, with the first panel reports the cross sectional dataset and second panel reports the panel dataset. All agglomeration variables are defined at the MSA level. A MSA contains a large population center and adjacent communities with a high degree of economic and social interactions. According to 2000 Census, about 76% of US population resides in MSAs. The variable *Population Size in the MSA* is the MSA's total number of all residence[1], which is a proxy for the size of local economy. Throughout the paper, the IPUMS person sampling weights (*PERWT*) are used in constructing demographic related agglomeration variables to ensure that they are representative at the national level. The variable *Total Employment in the MSA* is an aggregated measure based on the Mid-March Employment records from CBP data. It measures the total number of all employed workers in the local labor market. The variable *Industry diversification Index in the MSA* is an inverse form of the Herfindahl index based on the MSA's sectoral employment concentration. It is similar to the measure used by Henderson et al. (1995)[2]. Additionally, the variable *Labor Demand Shock in the MSA* is a plausible exogenous measure of the inter-temporal shifts of

---

[1] This refers to the general population in the MSA, includes all men and women, children and adults.

[2] The variable *Industry Diversification Index in the MSA* follows the method of Henderson et al. (1995), in which:
$Diversification\ Index_j = \sum_k (e_k/e)^2 / \sum_k (e_{jk}/e_j)^2$
The term e_{jk}/e_j is the j_{th} MSA's industry sector k_{th}'s share of employment in the total employment of the j_{th} MSA. Similarly, term e_k/e is the k_{th} sector's share of employment of the total employment at the national level. The higher value of this index suggests larger diversification in local industrial employment. The industry categories are defined by the first two digits of the 6-digit NAICS code. All establishments are counted into their respective two-digit industry categories, and there are 19 such categories defined in total.

MSA's labor demand. I construct this variable using the same method developed by Bartik (1991)[1].

3. Agglomeration and Inter-temporal Change of Hours Worked

Drawing on the labor pooling model developed by Krugman (1991), this section develops an empirical strategy using panel data analysis to test how labor pooling reduces the inter-temporal changes of average hours worked. Krugman (1991) argued that labor market pooling could be considered as a mechanism to reduce the employment and unemployment risk associated with firm-specific shocks. It helps spatially concentrated firms and workers to improve their abilities to adapt to firm-specific labor demand shocks, and benefit from more stable wages and hours worked. In this section, I present a version of Krugman's (1991) labor pooling model that relates to hours worked, and find that labor pooling in agglomeration economies reduces the inter-temporal changes of hours worked.

3.1 A labor market pooling model

Figure 1 illustrates a version of Krugman's (1991) model on labor pooling. Figure 1A shows an economy where a large number of firms are operating close to a large pool of skilled workers. In the presence of idiosyncratic firm-specific shocks, one firm's good times (positive shocks) may coincide another firm's bad times (negative shocks). A firm can easily hire more workers in good times and fire redundant workers in bad times, since there is a pool of workers readily available in the local labor market. Meanwhile, workers simply switch jobs between

---

[1] Where: $\textit{MSA Labor Demand Shock}_{t,t\text{-}1,MSA} =$
$$\sum\nolimits^{n}_{I=1}\{[(\textit{Employment}^{i}_{t,national}/\textit{Employment}^{i}_{t\text{-}1,national})\text{-}1] \times (\textit{Employment}^{i}_{t\text{-}1,MSA}/\textit{Employment}_{t\text{-}1,MSA})\}$$
It is the summary of all present industries' national level employment growth rate from time *t* to *t-1*, interact with each industry's share of the total local employment. The changes of industry's employment at the national level are plausibly uncorrelated with the local labor supply, but are correlated with the local labor demand shifts.

firms and always work for a stable wage rate and hours worked. This makes the firm-specific labor supply schedule to be completely elastic. In contrast to Figure 1A, Figure 1B shows a one-company town where all local skilled workers are employed by the company. Since there is no labor pooling and the number of available local skilled workers is limited, the labor supply schedule is inelastic in this case. When this company is experiencing labor demand shifts, it relies on workers' overtime work as a mechanism to adjust labor input. This indicates a greater degree in across-time volatility of hours worked. Additionally, Figure 1C shows an economy where the degree of labor pooling, proxy by the number of employers, is in between that of Figure 1A and Figure 1B. In this economy, labor pooling is not large enough to accommodate firm-specific labor demand shifts. Each firm thus needs to partially rely on overtime work to adjust its labor input. This situation results in an upward sloping firm-specific labor supply schedule.

Comparing across Figure 1A, 1B, and 1C, it is immediately clear that labor pooling tends to reduce inter-temporal change of hours worked. This pattern suggests that the average number of hours worked in an MSA tends to have less fluctuation across sample years, if there is a larger number of business establishments in that MSA. Drawing on the model presented above, I carry out a set of regression analysis to test the effects of labor pooling. Two different measures of inter-temporal change of hours worked were used as dependent variables in the regressions. The first measure is the *standard deviation of MSA's average hours worked* across sample years, and the second measure is the *first-differenced average of hours worked in each MSA*.

3.2 Results based on standard deviation of average hours worked

I start with a simple cross sectional analysis to test how labor pooling reduces the standard deviation of MSA's average hours worked across sample years. The data used in this analysis is an MSA-level cross sectional dataset derived from the MSA panel described earlier in Section 2. One of the variables provided in the MSA panel is the average hours worked in the MSA for each sample year. Using this variable, I calculate the standard deviation of average hours worked in the MSA across all sample years. In the regression, this standard deviation is used as the dependent variable. The explanatory variables are a set of MSA agglomeration characteristics. Each explanatory variable is generated by taking the averaged value (over all sample years) of the respective time-variant agglomeration variables in the MSA panel.

Table 2 reports the results from the cross sectional analysis. Column (1) shows the estimates based on the stratified subsample of skilled professional workers[1]. The key explanatory variable is the *Average Number of Establishments in the MSA*, used as a proxy for the degrees of labor pooling. The estimated elasticity on this variable is -19.5% and is statistically significant (with t-ratio of -1.69). It shows that the average hours worked tends to have smaller volatility in MSAs where a larger number of business establishments are clustered together enjoying a greater degree of labor pooling. This result is consistent with the predictions from Krugman's labor pooling model.

Column (2) of Table 2 reports results based on the stratified subsample for less-skilled nonprofessional workers. The estimated coefficient on variable *Average Number of Establishments in the MSA* is not statistically significant in this case. Comparing to the significant estimate on the same variable for professional workers in column (1), it suggests that

[1] See earlier descriptions on the stratification of professional workers and nonprofessional workers in Section two of this paper on Data and Variables.

the spillover from labor pooling perhaps does not affect the volatility of hours worked for the less-skilled workers.

Among the other agglomeration variables listed in Table 2, I use the variable *Average Population Size in the MSA* to control for the impact from the size of the local economy. The estimated elasticity on this variable is -36.3% for professionals in column (1), and is -24.9% for nonprofessionals in column (2). Both estimates are statistically significant. This result implies that the size of economy also tends to reduce volatility of hours worked for works regardless skill levels, when controlling for labor pooling and other agglomeration variables. This result is congruent with the central place theory, which suggests that a more diversified and vibrant economy may offer more stability in the local labor market.

3.3 Results based on first-differenced average hours worked

Next I carry out a set of panel data analysis using the MSA-level panel described earlier in Section 2. In these panel regressions, I use the first-differenced MSA average hours worked as the dependent variable, which all measure the change of the MSA average hours worked between any two consecutive sample years.

Table 3 reports results. I experiment with several different specifications in formatting the dependent variable in the terms of its absolute value. The reason to use absolute value of the fist-differenced average hours worked is to ensure that the dependent variable correctly measures the magnitude of its change between any two sample years, regardless whether the change is an increase or a decrease. In Table 3, column (1) and (2) report results using the logged form of absolute value of the first-differenced hours worked as the dependent variable, column (3) and (4) use absolute value of first-differenced hours worked as the dependent variable, and the last two columns (5) and (6) use absolute value of first-differenced log hours worked as the dependent variable. For each different dependent variable, the regression is carried out twice and separately for professional workers and nonprofessional workers. In all panel data regressions listed in Table 2, the MSA fixed effects are applied to account for any potentially unobserved time-invariant MSA characteristics.

Several patterns clearly stand out after inspecting the estimated coefficients on the variable of interest, number of establishments in MSA, across all six columns in Table 3. First, for all three different forms of dependent variable, the estimated coefficients for the professional workers are all negative and statistically significant (all above the 10% level). This result is consistent with the predictions from the labor pooling model, showing the spillover from labor pooling tends to reduce the inter-temporal change of average hours worked. Second, for nonprofessional workers, all three estimated coefficients on number of establishments in MSA are much smaller in magnitude relative to that of professional workers, although they do not have strong statistical significance (all below 10% level). This is congruent with the assumption embedded in the Krugman's (1991) labor pooling theory, which assumes that the spillover from labor pooling is only effective among skilled workers, but not among the less-skilled workers.

In all regressions reported in Table 3, the labor demand shock in MSA is included as an explanatory variable to control for the exogenous labor demand shifts in the MSA. In addition, an interaction term between the labor demand shock in MSA and the number of establishments in MSA is also included to control for any potential impact jointly determined by the two variables. The variable Industry diversification index in MSA controls for the industrial composition of sectorial employment in the MSA.

4. Agglomeration and Cross Sectional Variance of Hours Worked

This section focuses on how cross sectional variance of individual's hours worked in an MSA varies with degrees of labor pooling. To identify this relationship, I adopt the second approach briefly described earlier in the introduction section, and use the cross sectional data drawn from year 2000 Census and CBP to perform empirical tests. To start, I first present a modified model of labor pooling based on Krugman (1991), with a focus on the cross sectional perspective of the variance of hours worked.

4.1 Labor pooling model focusing on cross sectional variance

Figure 2 illustrates a modified version of labor pooling model and shows how cross sectional variance of hours worked would vary with degrees of labor pooling. This model is based on the stylized model developed by Krugman (1991) and is slightly different from the model presented in the previous section focusing on the inter-temporal change of hours worked. In the presence of idiosyncratic firm-specific shocks, one firm's good times are more likely to correlate with another firm's bad times, when a larger number of firms are nearby. Figure 2A shows an economy in which a large number of firms are clustered together, and the negative correlation between their demands for skilled labor prompt a greater degree of labor pooling. Since there are large pools of skilled labor, the supply schedule of the labor market is elastic with a relative flat slope. At any given moment, the firm (for example, firm 2) experiencing a negative shock would employ a small number of skilled workers. In the same time, another firm (for example, firm 1) experiencing a positive shock would employ a large number of skilled workers. Workers work for firm 1 and firm 2 receive similar wage and work for similar numbers of hours.

In contrast, Figure 2B shows an economy with a relative smaller number of firms. Since there are only a few firms in the economy, one firm's good times are less likely to coincide with another firm's bad times, resulting in a lesser degree of labor pooling. Therefore, the labor market supply schedule is inelastic and displays a relatively steeper slope. In this case, firms find it costly and difficult to vary the sizes of their workforces (through hiring and firing), and have to rely more on changing hours worked to acquire the optimal level of labor input. For example, at any given moment, the firm (firm 1 in Figure 2B) experiencing a positive shock would pay a higher wage to its skilled workers for overtime work. In the same time, the firm (firm 2 in Figure 2B) experiencing a negative shock would pay a lower wage asks its workers to work a reduce number of hours. As a result, there is a large variance in the wages and number of hours worked of workers employed by different firms.

By comparing between Figure 2A and 2B, it is apparent that the economy with greater degrees of labor pooling has less cross sectional variance of hours worked. The opposite is true for the economy with lesser degrees of labor pooling. This suggests that the cross sectional variance of hours worked tends to be smaller in the MSA with a larger number of business establishments. Next I carry out a set of regression analysis to test this relationship.

4.2 Main Results based on cross sectional variance of hours worked

To test how cross sectional variance of hours worked varies with degree of labor pooling, the first step is to construct a measure of variance of hours worked conditioned on worker's socio-demographic attributes. Different individuals may have different levels of hours worked, and this difference is determined not only by the characteristics of local economies (such as labor pooling), but also by workers' demographic attributes as well as their occupations and industries (Rosenthal and Strange, 2008).

I first regress individual worker's *Usual Hours Worked Per Week* on a set of standard socio-demographic variables, such as age, education, race, and others. Occupation and industry effects are included in the regressions to control for the unobserved differences across industries and occupations at the individual level. The dataset used is the cross sectional data from the sample year 2000. A detailed description of this dataset is provided in Section 2. This reduced form approach is similar to the labor supply model used by Rosenthal and Strange (2008). Table 4 reports the estimated coefficients, and they are consistent with the findings from those in Rosenthal and Strange (2008)[1].

Next, based on the estimated coefficients reported in Table 4, I generate the residuals of hours worked of each worker, and group them using their respective MSA codes. For each MSA, I calculate its standard deviations of the individual-level residuals of hours worked. This standard deviation is MSA-specific measure, representing the cross sectional variance of conditioned hours worked. It is used as the dependent variable in the following estimations.

Table 5 shows the estimations of agglomeration's impact on the cross sectional variance of hours worked. The dependent variable is the logged form of *standard deviation of residuals of hours worked in MSA*. The explanatory variables are a set of MSA-level agglomeration variables. For example, the variable *number of establishments in MSA* is a proxy for degrees of labor pooling. The variable *population size in the MSA* controls for the size of local economy.

[1] One thing worth noting is that this paper's labor supply model uses hours worked directly as dependent variable while previous studies (for example, Rosenthal and Strange, 2008) uses the log form of hours worked. Once converted to the log form, my estimates are consistent with previous studies.

The *industry diversification index in the MSA* controls for the industrial composition of sector employment in the MSA. The *total employment in the MSA* is a measure of the total size of local employed work force.

It is important to note that all standard errors reported in Table 5 are bootstrapped standard errors. Since the dependent variable *standard deviation of residuals of hours worked in MSA* is calculated based on the residuals generated from the regression on hours worked, the directly estimated standard errors become incorrect and are not valid for statistical inference[1]. To obtain correct standard errors for the estimates on the agglomeration variables, I apply a bootstrapping[2] procedure on both the first regression on hours worked and the second regression, and report the bootstrapped second stage coefficients and the associated bootstrapped standard errors in Table 5. Throughout this section, all two-step estimations on cross sectional variance of hours worked are bootstrapped for 999 replications to generate correct standard errors on agglomeration variables.

Column (1) in Table 5 reports the estimates based on the subsample for the skilled processional workers. Column (2) reports estimates based on the subsample for the less-skilled nonprofessional workers. The most notable estimate is the coefficients on the variable of interest, number of establishments in MSA. For both professional and nonprofessional workers, the coefficients are negative in value and statistically significant. This result is consistent with the predictions from the labor pooling model, that agglomeration reduces the cross sectional variance of hours worked through labor pooling.

---

[1] It is well known that the two-stage procedures can yield consistent estimates of the second stage parameters, but the directly estimated standard errors in the second stage regression are invalid for statistical inference. A detailed econometric theoretical analysis and proof can be found in Murphy and Topel (1985).

[2] See Efron and Tibshirani (1994) for a comprehensive study on bootstrapping. The bootstrapping method is a nonparametric procedure which does not require prior distribution assumptions on error terms, and can generate valid standard errors based on random re-sampling.

More importantly, there is a distinct difference between professional and nonprofessional workers in their estimated coefficients on variable *Number of establishments in MSA*. For professional workers in column (1), the estimated elasticity is -22.7% (with bootstrapped z-ratio of -2.19). This estimate is 12 percentage points larger than the estimated elasticity for nonprofessional workers in column (2), which is -11.4% (with bootstrapped z-ratio of -2.35). This difference indicates that the effect of labor pooling is stronger among skilled workers than those less-skilled workers. This pattern is consistent with the results in Sections 3 of this paper, and is also congruent with the skilled labor assumption in the labor pooling model.

The estimates on other explanatory agglomeration variables can shed further light on how other characteristics of local economy may affect the cross sectional variance of hours worked. For instance, the estimated coefficients on the variable *population size in MSA* are positive for both professional and nonprofessional workers, indicating that larger urbanized areas tends to have higher cross sectional variance in hours worked for both type of workers, with everything else equal. This is plausible because individuals may have heterogeneous tastes in how many hours they wish to spend on working, and a larger number of workers are likely to have a greater variety of preferences in hours worked. The positive sign of the estimated coefficients on *industrial diversification index in MSA* indicates that a more diversified industrial composition may leads to greater variance of hours worked.

4.3 Results based on MSA-stratified cross sectional variance of hours worked

One potential concern with the main results in the previous section is that the estimates on individual worker's attributes in the first-step regression are restricted to be the same across all MSAs. This can be problematic because the impact of worker's attributes on their hours worked may correlated with his location. Keeping their estimates fixed across MSAs may result in potential heteroskedasticity in the error terms. Consequently, the predicted residuals may be correlated within MSA clusters, causing biased estimates on the agglomeration variables in the regression on standard deviations of hours worked.

However, it is not completely unreasonable to restrict the coefficients in the hours worked regression to be fixed for all MSAs. Perhaps some of the MSA-specific agglomeration characteristics are associated to certain attributes shared by the workers in the MSA. In my main specifications, these agglomeration variables are estimated in the second step MSA-level regression, but are omitted from the first step individual-level hours worked regression. If the coefficients in the first step individual-level hours worked regression are restricted to be the same for all MSAs, the effect of these omitted MSA agglomeration characteristics will be captured by the generated residuals of hours worked. When the MSA-specific variance of these residuals is used as the dependent variable in the second step regression, the effect of these MSA characteristics could be properly estimated. Therefore, restricting the coefficients in hours worked regression across MSAs may help keep some valuable information about the MSA agglomeration characteristics in the residuals of hours worked, thus improve the estimates on the agglomeration variables in the second step regression on the variance of residuals of hours worked.

Column (3) and column (4) in Table 5 report estimations using an alternative specification to address the concerns described above. More specifically, the first step individual-level hours worked regression, from which the residuals are generated, is MSA-stratified. The coefficients on worker's attributes are estimated separately for each MSA, allowing them to vary across MSAs. Using these MSA-stratified estimates, I generate a new set of residuals of hours worked from the first-step regression. Next I follow the same steps from the main specification, calculating the MSA-specific standard deviations of hours worked, and use it as the dependent variable to regress on the same set of agglomeration variables in the second-step regression.

Column (3) presents results for professional workers. The estimated elasticity with respect to the *number of establishments in the MSA* is -21.4%, which is statistically significant at 10% level. This is consistent with the estimates from column (1) of Table 5. It is also congruent to my prediction that agglomeration tends to reduce cross sectional variance of hours worked. Column (4) reports results for nonprofessional workers. The estimated coefficient on *number of establishments in the MSA* is -8.3% but does not have strong statistical significance (with a z-ratio of -1.51). This estimate is 13 percentage points lower in magnitude than that of the professional workers in column (3). The distinct difference between professional and nonprofessional workers is similar with the pattern found in column (1) and (2). It is also consistent with the assumption that the spillover from labor pooling is only effective among skilled workers.

4.4 Robustness Tests

This subsection investigates the degree to which the estimates from previous subsections (reported in Table 5) are robust to alternative geographic measures for agglomeration economies, and to data from different sample year. The cross sectional data used in previous estimations is merged by Census and CBP data for sample year 2000, and the geographic area of an agglomeration economy is defined at the MSA level. In this subsection, I experiment the same specifications on two alternative datasets in cross sectional setting.

Table 6 reports estimation results based on the first alternative dataset. In this dataset I use WPUMA[1] to define the agglomeration economy's geographic area, and merge the 2000 census data with the 2000 CBP data at the WPUMA level. The estimates reported in Table 6 are consistent with my main results in earlier sections. In column (1) for professional workers, the estimated elasticity with respect to the number of establishments in the WPUMA is -21% and statically significant (with z-ratio of -4.67). Column (2) reports that nonprofessional workers' estimated elasticity of the same variable is -19% and is also statistically significant (with z-ratio of -8.76). These estimates show that the number of establishments in the WPUMA has a negative relationship with the cross sectional variance of hours worked, and the magnitude of this impact is larger for skilled workers relative to the less skilled workers. These results are consistent with the estimates obtained by using the MSA-level cross sectional data (reported in Table 5). This suggests a strong robustness of my results subject to different geographic measures.

Table 7 presents estimations based on the second alternative cross sectional data from sample year 2008, and the estimates are similar with the results based on data from sample year 2000. This 2008 data is merged by the single year ACS data and the CBP data from 2008, and is matched at MSA level. The 2008 cross sectional dataset has the exact same parameters as the cross sectional data of sample year 2000. Due to the ACS survey design, however the sample

[1] PUMA is short for Public Use Microdata Areas, a geographic statistic unit used by the U.S. Census Bureau. WPUMA (Work PUMA) is identified by the first 3 digits of the full 5 digits PUMA code, it includes all PUMA areas within. Prior evidence suggested that agglomeration economies attenuate rapidly (Rosenthal and Strange, 2003, 2005), and since workers generally live and commute to work inside a work PUMA, I consider work PUMA as a local labor market with a pooled labor force that can be shared by locally clustered firms. In total, there are 1,153 work PUMAs identified in our sample.

size of the 2008 data is much smaller than that of the 2000 Census sample[1]. All reported coefficients in Table 7 are estimated using the same methods from the main specifications.

Column (1) in Table 7 reports results for professional workers. The estimated coefficient on *number of establishments in the MSA* is -22.9, and its bootstrapped standard error fails to show statistical significance. This weak significance could be attributed to the small sample size of the 2008 ACS data, which could potentially reduce the dataset's statistical power. Despite the low statistical significance, this estimate can still shed some light on robustness of earlier results. The negative sign of this coefficient again confirms the findings from the earlier estimation, that the cross sectional variance of hours worked tends to be reduced by the labor market pooling. Furthermore, this estimated elasticity of the professional workers is three percentage points larger relative to the nonprofessional workers' estimates reported in column (2). This relative difference between the skilled workers and the less skilled workers is consistent with the same pattern I find based on the cross sectional data of sample year 2000.

5. Conclusions

This paper is the first one to systematically document the relationship between agglomeration and the volatility of hours worked. The paper provides empirical evidence in support of the labor pooling model developed by Krugman (1991), and show that the spatial concentration of firms and labor reduces the volatility of hours worked associated with idiosyncratic firm-specific shocks. My test results are robust to different measures of volatility of hours worked, as well as to data from different sample years. Moreover, my findings also

[1] For the professional group, our customized sample from the ACS data is 26% of the size of the customized sample from 2000 Census 5% data. For nonprofessional group, the ACS 2008 customized sample size is 20% of the size of the 2000 Census 5% customized sample.

indicate that this spillover effect of labor pooling is more significant and stronger among skilled professional workers, relative to the less skilled nonprofessional workers.

This paper adds to a small but important literature that provides evidence on the microeconomic foundations of agglomeration economies. Labor market pooling is one of the many potential sources of agglomeration economies. The previous literature has offered many theories in explaining the micro foundations of labor market pooling, but has had less success in offering empirical evidence to identify the specific mechanisms. This paper provides direct evidence to a specific type of spillover from labor pooling, showing how labor pooling helps the spatially concentrated firms smooth labor market fluctuations and reduce firm-specific employment and unemployment risk. This paper is one of very few studies to have provided empirical evidence to identify the micro foundation of labor market pooling, and to help explain why cities are productive places.

References

Bartik, Timothy J., 1991. "Who Benefits From State and Local Economic Development Policies." W.E. Upjohn Institute for Employment Research: Kalamazoo, MI.

Combes, Pierre Philippe, and Duranton, Gilles, et al., 2007. "Estimating agglomeration effects: Does playing with different instruments give a consistent tune?" mimeo

Duranton, Gilles, and Puga, Diego, 2004. "Microfoundations of Urban Agglomeration Economies," in Henderson, J.V., and J. F. Thisse (Eds.): *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam: Elsevier, 2063-2117.

Efron, B., and Tibshirani, R. J., 1994. "An Introduction to the Bootstrap," Boca Raton, FL: CRC Press.

Gerlach, Heiko, Rønde, Thomas, and Stahlg, Konrad, 2009. "Labor pooling in R&D intensive industries," *Journal of Urban Economics* 65, 99–111.

Helsley, R.W., and Strange, W.C., 1990. "Matching and Agglomeration Economies in a System of Cities," *Regional Science and Urban Economics* 20, 189-212.

Overman, Henry G., and Puga, Diego, 2010. "Labor Pooling as a Source of Agglomeration: An Empirical Investigation," *Agglomeration Economics*, University of Chicago Press, 133 - 150

Henderson, I.V., Kuncoro, A., and Turner, M., 1995. "Industrial development in cities," *Journal of Political Economy* 103, 1067-1085

Krugman, Paul. R., 1991. "Geography and trade," Cambridge, MA: MIT Press.

Marshall, Alfred., 1920, *Principles of Economics*, London: MacMillan.

Murphy, Kevin M, and Topel, Robert H, 2002. "Estimation and Inference in Two-Step Econometric Models," *Journal of Business & Economic Statistics*, American Statistical Association, vol. 20(1), pages 88-97, January.

Rosenthal, Stuart S. and Strange, William C., 2004. "Evidence on the Nature and Sources of Agglomeration Economies," in Henderson, J.V., and J. F. Thisse (Eds.): *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam: Elsevier, 2119–2172

Rosenthal, Stuart S. and Strange, William C., 2008. "Agglomeration and Hours Worked," *The Review of Economics and Statistics*, MIT Press, vol. 90(1), pages 105-118, November.

Ruggles, Steven, Matthew Sobek, et al., Integrated Public Use Microdata Series: Version 2.0, Minneapolis: Historical Census Projects, University of Minnesota, http://www.ipums.org.

TABLE 1

Descriptive Statistics of Agglomeration Variables at MSA-level[a]

	MSA-level Cross Sectional Data (Sample Year 2000)			
	Mean	Std. Dev.	Min.	Max.
Population size in the MSA	770,347	1,590,788	100,506	17,200,000
Industry Diversification Index in the MSA	0.89	0.13	0.39	1.16
Total Employment in the MSA	337,888	700,821	26,236	7,266,626
No. of Establishments in the MSA (>50 employees Est.)	1,190	2,435	91	23,878
No. of White Male Full-time Professional Workers in MSA	12,941	32,445	464	368,769
No. of White Male Full-time Nonprofessional Workers in MSA	43,168	87,382	3,748	905,257
	MSA-level Panel (From Sample Year 2005 to 2009)			
	Mean	Std. Dev.	Min.	Max.
Population size in the MSA	813,499	1,651,158	98,329	17,800,000
Industry Diversification Index in the MSA	0.83	0.12	0.28	1.07
Total Employment in the MSA	348,010	704,350	29,395	7,329,855
No. of Establishments in the MSA (>50 employees Est.)	1,237	2,450	100	23,867
No. of White Male Full-time Professional Workers in MSA	62,516	72,975	67	273,917
No. of White Male Full-time Nonprofessional Workers in MSA	99,949	115,650	1,138	475,767

a. All MSA-level characteristics are based on the sample combined by the male full-year workers data from IPUMS ACS and Census data and the business establishment data from CBP. There are total 275 MSAs identified in both the cross sectional data and the panel data.

TABLE 2

Volatility of Average Hours Worked at the MSA-level Across Sample Years (From 2005 to 2009)

Dependent Variable: Log Std. Dev. of MSA's Average Hours Worked From 2005 to 2009[a] (t-ratios in parentheses)

	Professional[a]	Nonprofessional[a]
	(1)	(2)
Log Avg Number of Establishments in the MSA 2005-2009	-0.20	-0.04
	(-1.69)	(-0.43)
Log Avg Std Dev of Labor Demand Shock[b] in the MSA 2005-2009	-0.06	0.57
	(-0.24)	(2.86)
Log Avg Population Size in the MSA 2005-2009	-0.36	-0.25
	(-2.99)	(-2.58)
Log Avg Industry Diversification Index[c] in the MSA 2005-2009	-0.02	0.18
	(-0.11)	(0.99)
Constant	6.25	6.25
	(4.12)	(5.18)
No. of Observations	276	276
Adj R squared	0.62	0.36
Root MSE	0.47	0.37

a. The Log Std. Dev. of MSA Yearly Average Hours Worked is defined as $ln\{square\ root[(\sum^{5}_{t=1}(H_t - H\ ))/(5-1)]\}$

b. Professional workers belong to management and professional occupations and have a master's or higher degree. Nonprofessional workers belong to nonprofessional and service occupations and are high school graduated or less.

c. The MSA Demand Shock is defined as the summation of each industry's employment growth rate at the national level, weighted by the industry's employment share in each MSA:

$$MSA\ Labor\ Demand\ Shock_{t,t\text{-}1,MSA} = \sum^{n}_{I=1}\{[(Employment^{i}_{t,national}/Employment^{i}_{t\text{-}1,national})\text{-}1] \times (Employment^{i}_{t\text{-}1,MSA}/Employment_{t\text{-}1,MSA})\}$$

For a more detailed explanation on this variable, see page 11.

d. The MSA Industry Diversification Index is defined as the inverted Herfindahl index: $\sum_k(e_{jk}/e_j)^2/\sum_k(e_k/e)^2$

TABLE 3
Volatility of MSA Average Hours Worked Across Sample Years (From 2005 to 2009):
Estimates Based on MSA-level Panel Dataset following MSAs From 2005 to 2009 (in parentheses: t-ratios based on robust standard errors)

Dependent Variable[a]	Log Absolute Value of First-Differenced MSA Avg Hours Worked $\ln\lvert H_{t,MSA} - H_{t-1,MSA}\rvert$		Absolute Value of First-Differenced MSA Avg Hours Worked $\lvert H_{t,MSA} - H_{t-1,MSA}\rvert$		Absolute Value of First-Differenced Log MSA Avg Hours Worked $\lvert\ln(H_{t,MSA}/H_{t-1,MSA})\rvert$	
	Professional[b]	Nonprofessional[b]	Professional	Nonprofessional	Professional	Nonprofessional
	(1)	(2)	(3)	(4)	(5)	(6)
Log Number of Establishments in MSA_t	-1.74	-0.04	-5.14	-1.40	-0.10	-0.03
	(-1.84)	(-0.05)	(-1.95)	(-1.14)	(-1.81)	(-1.06)
Labor Demand Shock in $MSA_{t,t-1}$[c]	-18.79	32.52	-20.46	14.42	-0.56	0.29
	(-1.03)	(1.9)	(-0.55)	(0.72)	(-0.72)	(0.65)
Log Number of Establishments in MSA_t × *Labor Demand Shock in* $MSA_{t,t-1}$[c]	2.20	-5.54	1.82	-3.37	0.06	-0.07
	(0.8)	(-2.13)	(0.36)	(-1.2)	(0.53)	(-1.17)
Industry Diversification Index in MSA_t[d]	2.29	2.00	2.86	2.59	0.05	0.06
	(1.1)	(0.91)	(0.54)	(1.02)	(0.48)	(1.09)
Constant	8.29	-1.12	29.02	7.37	0.55	0.15
	(1.58)	(-0.23)	(2.06)	(1.11)	(1.92)	(1.00)
No. of Observations	1104	1104	1104	1104	1104	1104
MSA Fixed Effects	276	276	276	276	276	276
R squared	0.50	0.38	0.47	0.42	0.48	0.41
Root MSE	1.02	1.06	2.36	1.17	0.05	0.03

a. "H" indicates yearly MSA average hours worked, t ϵ [1,5], MSA ϵ [1, 276]
b. Professional workers belong to management and professional occupations and have a master's or higher degree. Nonprofessional workers belong to nonprofessional and service occupations and are high school graduated or less.
c. The Labor Demand Shock in MSA is defined as the sum of each industry's employment growth rate at the national level, weighted by the industry's employment share in each MSA:

$$MSA\ Labor\ Demand\ Shock_{t,t-1,MSA} = \sum_{I=1}^{n} \{[(Employment^{i}_{t,national}/Employment^{i}_{t-1,national})-1] \times (Employment^{i}_{t-1,MSA}/Employment_{t-1,MSA})\}$$

For a more detailed explanation on this variable, see page 11.
d. The variable MSA Industry Diversification is defined as the inverted Herfindahl index: $\sum_k(e_{jk}/e_j)^2/\sum_k(e_k/e)^2$

TABLE 4
Hours Worked Regression at Worker-level: Estimates Based on Data From Sample Year 2000
Dependent Variable: Usual Hours Worked Per Week (t-ratio in parentheses)

	Professional[a] (1)	Nonprofessional[a] (2)
PhD degree[b]	2.57	-
	(20.21)	-
Professional degree[b]	2.68	-
	(22.53)	-
High school graduate[c]	-	0.47
	-	(8.21)
Have children younger than 10	-0.11	0.36
	(-1.58)	(7.46)
Age	0.23	0.28
	(5.98)	(20.02)
Age squared	-0.01	-0.00
	(-9.81)	(-20.92)
Married	1.17	1.36
	(13.41)	(27.19)
Black	-1.47	-1.54
	(-9.11)	(-16.24)
Asian	-1.64	0.04
	(-11.86)	(0.13)
Hispanic	-2.01	-0.84
	(-3.68)	(-8.33)
Other race	-0.43	-0.26
	(-0.83)	(-1.18)
Immigrated 6-10 years ago[d]	0.53	-0.02
	(1.68)	(-0.12)
Immigrated 11-15 years ago[d]	1.17	0.04
	(4.54)	(0.26)
Immigrated 16-20 years ago[d]	1.84	-0.07
	(5.19)	(-0.32)
Immigrated >21 years or US citizen[d]	1.85	0.07
	(7.3)	(0.42)
Speak English	0.47	0.49
	(1.01)	(3.01)
Log commute time	0.01	-0.06
	(0.19)	(-1.6)
Constant	44.44	41.08
	(31.13)	(52.93)
No. of Occupation effects	8	8
No. of Industry effects	19	19
No. of observations (individuals)	150,048	459,531
R squared	0.07	0.05
Root MSE	10.62	9.29
No. of MSAs (clusters adjusted by Std. Err.)	275	275

a. Professional workers belong to management and professional occupations and have a master's or higher degree. Nonprofessional workers belong to nonprofessional and service occupations and are high school graduated or less.
b. Omitted categories for professional workers are master's degree.
c. Omitted categories for nonprofessional workers are not high school graduated.
d. Omitted category is immigrated in the last five years.

TABLE 5

Cross Sectional Variation of Hours Worked in MSA: Estimates Based on Cross Sectional Data From Sample Year 2000

Dependent Variable: Log of Std. Dev. of Residuals of Hours Worked in MSA[a] (bootstrapped t-ratio in parentheses)

	MSA Level Estimates Based on Unstratified Estimates from Hours Worked Regression Reported in Table 4[b]		MSA Level Estimates Based on MSA-Stratified Estimates from Hours Worked Regression (not reported)[c]	
	Professional[d]	Nonprofessional[d]	Professional	Nonprofessional
	(1)	(2)	(3)	(4)
Log Population Size in the MSA	0.07	0.11	0.30	0.09
	(2.14)	(2.21)	(2.34)	(1.57)
Industry Diversification Index in the MSA	0.09	0.33	0.13	0.24
	(1.43)	(11.71)	(1.5)	(6.63)
Log Total Employment in the MSA	0.16	0.02	0.22	-0.01
	(1.52)	(0.29)	(1.63)	(-0.02)
Log Number of Establishments in the MSA	-0.23	-0.11	-0.21	-0.08
	(-2.19)	(-2.35)	(-1.72)	(-1.51)
Constant	0.98	1.46	-0.17	1.36
	(1.61)	(5.13)	(-0.22)	(4.17)
No. of observations (MSAs)	275	275	275	275
Adjusted R squared	0.05	0.30	0.24	0.27
Root MSE	0.13	0.08	0.16	0.07

a. The Dependent Variable for the 1st stage is *Usual Hours Worked Per Week* for all two-stage procedures. All demographic variables listed in Table 4 are also included in each 1st stage regressions in all two-stage procedures, but their coefficients are suppressed in this table to conserve space.

b. Both hours worked regression showed in Table 4 and the standard deviation regression showed in this table are bootstrapped for 999 reps. I only report the regression estimates on agglomeration variables here in this table.

c. In this MSA-stratified procedure, the hours worked regressions (reported in Table 4) are stratified by MSA. There are in total 275 hours worked regressions. Each regression corresponds to a specific MSA. This entire procedure is also bootstrapped for 999 reps.

d. Professional workers belong to management and professional occupations and have a master's or higher degree. Nonprofessional workers belong to nonprofessional and service occupations and are high school graduated or less.

TABLE 6
Cross Sectional Variation of Hours Worked in WPUMA: Estimates Based on Data From Sample Year 2000[a]
Dependent Variable: Log of Std. Dev. of Residuals of Hours Worked (bootstrapped t-ratio in parentheses)

	WPUMA Level Estimates Based on Unstratified Estimates From Hours Worked Regression[bc]	
	Professional[d] (1)	Nonprofessional[d] (2)
Log Population size in the WPUMA	0.20	0.16
	(4.24)	(7.36)
Industry Diversification Index in the WPUMA	0.07	0.31
	(2.12)	(20.09)
Log Total Employment in the WPUMA	0.19	0.17
	(4.09)	(7.63)
Log Number of Establishments in the WPUMA	-0.21	-0.19
	(-4.67)	(-8.76)
Constant	1.23	1.19
	(4.44)	(9.2)
No. of observations (WPUMAs)	1,150	1,151
Adjusted R squared	0.05	0.22
Root MSE	0.15	0.10

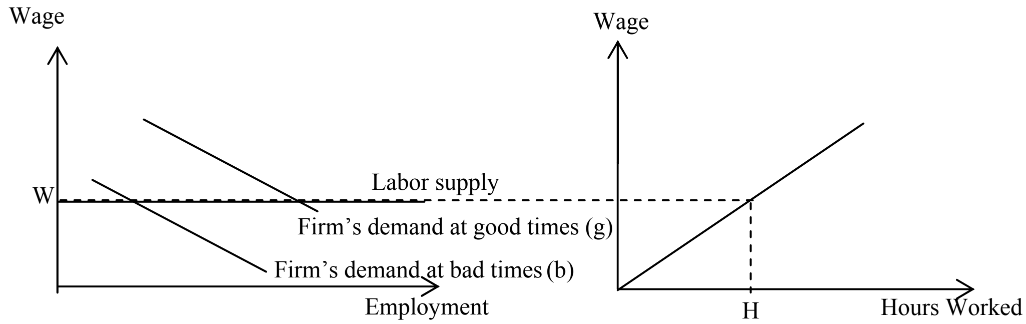
a. All the workers demographic variables listed in table 4 are also included in the hours worked regression of this procedure. Their coefficients are suppressed in this table to conserve space.
b. The entire estimation procedure is Bootstrapped for 999 replications. The bootstrapping z-ratios are reported in parentheses.
c. Similar results are estimated by a procedure with WPUMA-stratified hours worked regressions.
d. Professional workers belong to management and professional occupations and have a master's or higher degree. Nonprofessional workers belong to nonprofessional and service occupations and are high school graduated or less.

TABLE 7
Cross Sectional Variation of Hours Worked in MSA: Based on The Cross Sectional Data From Sample Year 2008[a]
Dependent Variable: Log of Std. Dev. of Residuals of Hours Worked (bootstrapped t-ratio in parentheses)

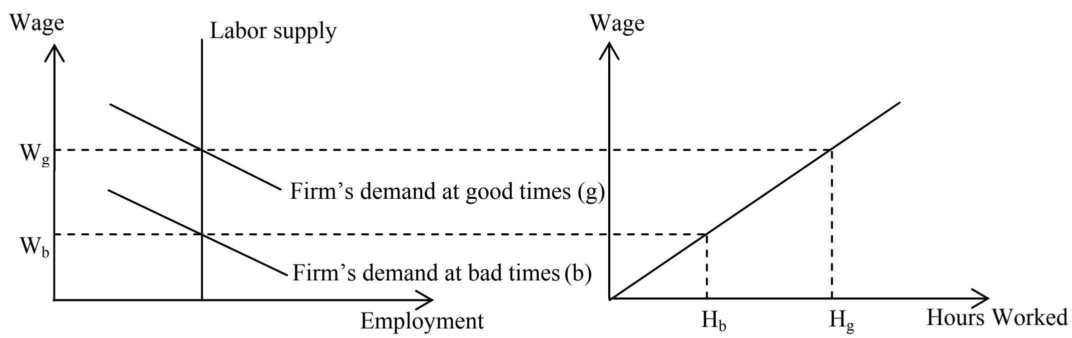
	MSA Level Estimates Based on Unstratified Estimates From Hours Worked Regression[b]	
	Professional[c]	Nonprofessional[c]
	(1)	(2)
Log Population size in the MSA	0.26	0.17
	(1.46)	(1.47)
Industry Diversification index in the MSA	0.05	0.23
	(0.28)	(2.62)
Log Total Employment in the MSA	0.21	0.21
	(1.16)	(1.78)
Log Number of Establishments in the MSA	-0.23	-0.19
	(-1.34)	(-1.65)
Constant	0.55	1.19
	(0.52)	(1.78)
No. of Observations (MSAs)	275	275
Adjusted R squared	0.02	0.03
Root MSE	0.22	0.14

a. This alternative sample is drawn from the IPUMS ACS 2008 and CBP 2008 data. The sample is restricted to male workers identified in MSAs, employed in business sectors and who have full-year employment status. The full-year status is defined as worked more than 50 weeks last year. All the workers demographic variables listed in table 3 are also included in the 1st stage regression of this two-stage procedure. Their coefficients are suppressed in this table to conserve space.
b. The entire estimation procedure is bootstrapped for 999 reps. The bootstrapping z-ratios are reported in parentheses.
c. Professional workers belong to management and professional occupations and have a master's or higher degree. Nonprofessional workers belong to nonprofessional and service occupations and are high school graduated or less.

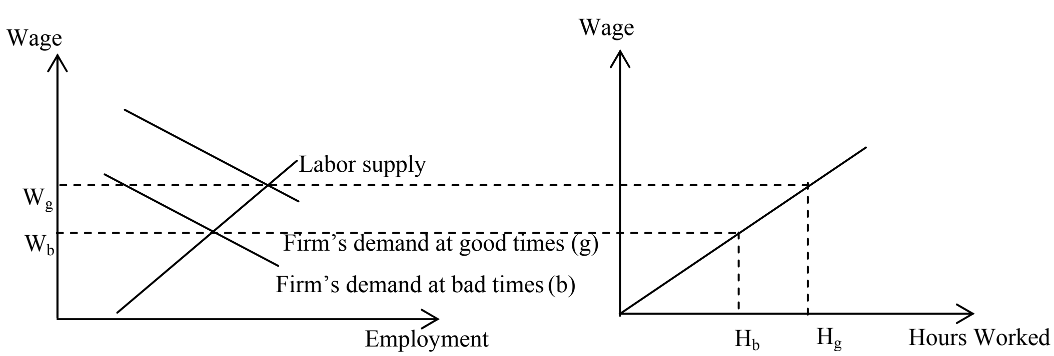
Figure 1: Labor pooling and inter-temporal change of hours worked



A: A firm operating close to large pools of skilled labor

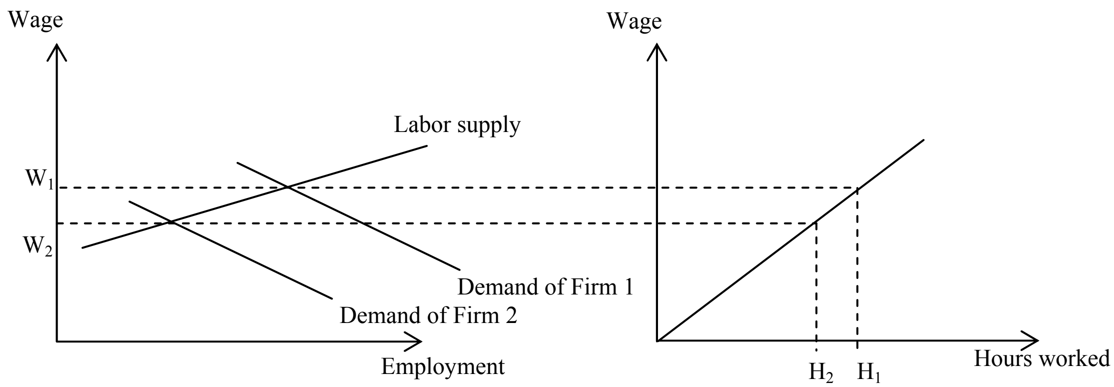


B: A firm operating in a one-company town with no pooling

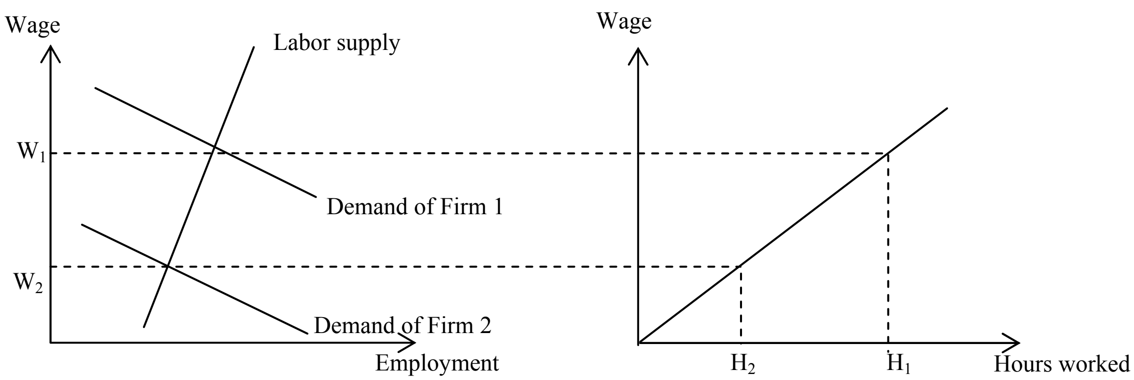


C: A firm operating close to middle-sized pools of skilled labor

Figure 2: Labor pooling and cross sectional variance of hours worked



A: A large number of firms operating close to large pools of skilled labor



B: A small number of firms operating close to small pools of skilled labor

Essay II: Homeownership and Men's Labor Supply: An IV Approach

1. Introduction

Homeownership is strongly associated with men's labor supply. A probit estimate suggests that for male household heads in the U.S., homeowners are 2.6% more likely to be employed relative to renters[1]. This positive effect, however, is potentially biased by the endogeneity of homeownership. On one hand, individuals with good jobs have better access to mortgage markets and are more likely to buy homes. On the other hand, homeowners have strong motivations to work in order to afford mortgage payments and homeownership costs. A number of studies (for example, Coulson and Fisher, 2008, Munch et al., 2006, Bottazzi, 2007, Del Boca and Lusardi, 2003)[2] have investigated the relationship between homeownership and labor supply. However, their results are mixed, and the endogeneity of homeownership remains as a challenging identification problem.

In this paper, I use a new set of instrumental variables (IV) to address this endogeneity problem, and to estimate the causal effect of homeownership on men's labor supply. The instruments are four dummy variables (whether the household has three children, four children, five children, and six or more children, respectively), which measures the size of single family households. I show that these family size instruments are highly correlated with homeownership, and the IV diagnostic test results suggest that my instruments meet all conventional requirements of a valid IV estimation.

I find that family size is closely related to a household's decision on whether to owner-occupy its homes. Using household-level data from the American Community Survey (ACS)

[1] See column (1) in Table 4, the estimated marginal effect on *Homeownership*. This probit estimation is based on a sample of married white male household heads of single-family households in the U.S., more information about the sample is provided in section 3.

[2] See Section 2 of this paper for a brief review of the previous studies on the relationship between homeownership and labor supply.

2010 3% file, I find that the likelihood of single family households being homeowners is positively associated with the presence of the first two children, but negatively associated with additional children beyond the first two kids. This result is consistent with the prediction from a conceptual model. In the model, homeownership is jointly determined by the household's consumption demand and investment demand for housing stocks. Both demands are assumed to be functions of the number of children in the household.

Drawing from the evidence that the third and additional children are highly correlated with homeownership, I use them as family size instruments to address the endogenous relationship between homeownership and men's labor supply. The IV estimates suggest that male household heads who own homes are 1.2% more likely to be employed relative to renters, after controlled for demographic and socio-economic attributes. Comparison of IV and probit estimates suggest that there is substantial upward endogeneity bias in the conventional cross-sectional estimates of the effect of homeownership on men's labor supply. In particular, the IV results are congruent with previous theories arguing that homeownership has a positive effect on labor supply, but do not support the Oswald hypothesis - that homeownership decreases employment.

I also examine the extent to which the effect of homeownership on men's labor supply differs across age groups. Using the instruments described above, I separately estimate IV models for men between ages 18 and 33, and for men between ages 34 and 55. I find that homeownership has a positive and significant effect on labor supply among younger men but it does not have a significant effect for men in older age. This result suggests that younger male homeowners have more incentive to work relative to older ones. In addition, I explore how family size relate to men's hours worked. OLS estimates suggest that family size has a significant and nonlinear incremental effect on men's hours worked. The first two children have positive effects but the third and additional children have negative effects.[3]

This paper makes three contributions. Firstly, using household-level dataset, the paper provides micro-data evidence on the positive effect of homeownership on men's labor supply. This evidence adds to a small but important literature on the spillover effects of homeownership. Secondly, I develop a new set of instruments to address the endogeneity of homeownership in estimating its impact on men's labor supply. Thirdly, this paper shows that the relationship between housing tenure status and family size is highly nonlinear and nonmonotonic.

The rest of paper is organized in the following manner. Section 2 provides some background on previous research. Section 3 presents the data and variables. Section 4 covers the effect of children on homeownership. Section 5 covers estimations of impact of homeownership on men's labor supply. Section 6 explores the effects on hours worked. Section 7 concludes.

2. Background

On the topic of homeownership's labor market impacts, the previous literature provides many competing theories which predict different effects, either positive or negative. There are several reasons to argue that homeownership has a positive impact on household's labor market outcomes. Firstly, a number of studies argue that homeowners are more likely to be employed relative to renters. Being a homeowner often means that the household has to bear additional financial responsibilities (such as mortgage repayments and maintenance costs) over extensive

[3] Lundberg and Rose (2002) argue that the negative incremental effects by children in addition to the first two kids are potentially caused by the heterogeneity bias in OLS estimates. After using a fixed-effects approach, they show that children in addition to the first two kids do not have significant increment effects on men's hours worked and hourly wages.

periods of time.[4] To fulfill the repayments of debts, homeowners are more likely to increase their labor supply and to work harder (for longer hours, or take a second job) relative to renters. Fortin (1995), Del Boca and Lusardi (2003), Bottazzi (2007), and Bottazzi et al. (2007) find that women from households with greater mortgage commitments are more likely to work, and tend to work longer. Flatau et al. (2003) and Munch et al. (2006) find homeowners tend to have shorter unemployment spells than renters, based on Australian and Danish data, respectively.

Secondly, several studies show that homeowners may have higher productivity and higher wage level relative to renters. Owner-occupiers on average tend to have longer residence spells, which not only increase their stability and social capital but also lead homeowners to focus more on work thus they become more productive (Rohe and Stewart, 1996; Dipasquale and Glaeser, 1999). Coulson and Fisher (2002) find that homeowners have higher wages relative to renters in the U.S. Munch et al. (2008) find similar result by using data from Denmark, and argue that the immobility of homeowners may cause them to invest more in local jobs, increasing their firm-specific productivity.

There are also reasons to believe that homeownership has a negative impact on households' labor market outcomes. For example, relative to renters, owner-occupiers tend to have lower mobility due to high costs of buying and selling homes (Ferreira, et al., 2010, and others[5]). Low household mobility, argued by Oswald (1997) and others, links homeownership to inferior labor market outcomes. Less-mobile homeowners are either constrained to their residential locations or tend to have longer commutes, thus they are less efficient in exploiting

[4] About 83% of single-family households in the Unites States who own their homes have mortgage debts, according to the author's calculation using a sample of white male household heads drawn from the ACS 2010 3% data file.
[5] For instance, Quigley (1987), Stein (1995), Genesove and Mayer (1997, and 2001), Chan (2001), and Engelhardt (2003).

job opportunities elsewhere than mobile renters. Other models based on search theories (Dohmen, 2005; Munch et al., 2006; Coulson and Fisher, 2008) further argue that this immobility of owner-occupiers constrains their ability to move and to accept job offers from firms in other locations. Therefore, the immobility exacerbates the spatial mismatch between jobs and workers' residential locations.

A number of previous studies offered evidence on the negative effect of homeownership by investigating the aggregate-level relationship between homeownership rates and unemployment rates. For example, Oswald (1997, 1999) and others[6] found positive correlation between homeownership rates and unemployment rates among OECD countries. However, this type of cross-county analysis usually suffers from potential problems of omitted variables and measurement errors.[7]

With respect to the relationship between homeownership and family size, many researchers have argued that households with large family size are more likely to be homeowners. It is reasonable to think that if there are more children in a single family household, the family would consume more housing and prefer more control over their living spaces. Therefore, they are more likely to become owner-occupiers relative to households with no child or small number of children. However, previous literature offered limited evidence to support this speculated relationship. A number of studies consider the inter-dependence between family's fertility

[6] Similar results based on aggregate macro data are also found by Green and Hendershott (2001), Partridge and Rickman (1997), and Pehkonen (1999). However, Barrios and Rodríguez (2004) found the correlation to be in the opposite direction.
[7] Many of the important individual-level attributes relate to homeownership labor market outcomes could not be properly accounted for by the aggregate measures.

decisions and their housing tenure status, but most of them focus on the birth of a new child (i.e., the first child)[8] rather than the total size of the household.

In the labor literature, the traditional focus of impacts of bearing children is centered on women's labor market outcomes, but several studies find that children also affect men's labor supply and earnings. For example, Lundberg and Rose (2002) show that the effects of children on the father's labor supply are highly nonlinear and nonmonotonic. Similar result is offered by Angrist and Evans (1998), who use gender differences of the first two children to predict the likelihood of the parents having a third child, and use the same-sex instruments to estimate the effect of the third child on women's and men's labor supply.

In the labor economics literature, the studies on the effect of children on men's labor supply is relative less extensive comparing to research on the effect of fertility on women's labor market outcomes. Women are traditionally viewed as the primary caretakers for children. However, it is reasonable to think that both the husband and the wife bear part of the time and financial cost (although in different proportions) for raising children. Assuming the labor market decisions of husbands and wives are inter-dependent, children should also affect men's labor supply, perhaps to a different extent from that of women's.

Several studies have examined the relationship between children and men's labor supply. For example, Lundberg and Rose (2002) find that the effect of children is highly nonlinear and nonmonotonic. They show that the first two children have positive and significant impact on

[8] For example, Mudler and Wagner (2001), Feijten and Mudler (2002), Courgeau and Lelievre (1992), Murphy and Sullivan (1985) all considers some aspects of the interaction between household housing types and birth of the first child. One possible exception is Kulu and Vikat (2007), who considers the relationship between selective moves into owned housing and birth of the first, second, and third children separately.

men's probability of being employed. Agrist and Evans (1998) find similar result, using gender differences of the two first-borns as instrument for the birth of a third child.

3. Data and Variables

The primary data of this study is a sample of married white male household heads from the American Community Survey (ACS) 2010 3% household-level data file. It is obtained from the IPUMS[9] database. For the purpose of examining the impact on men's labor supply, I restrict my sample to include married white male heads of single family households. A typical single family household consists of a married couple and their offspring.

Following the approach used by Angrist and Evans (1998) as well as Coulson and Fisher (2008), I limit the male household heads sample to those who are aged between 18 and 55, and whose wives are aged between 18 and 35. The restriction on wife's age is to avoid generating a selective sample. There are several reasons. Firstly, I only use the number of children below age 18 to proxy for family size. A child over age 18 is more likely to live in another residence (i.e. in college dorm), and have ambiguous effects on its parents' housing demand and labor supply. Secondly, since ACS does not track children who are currently not living with their parents, the 18 cut-off on children's age will ignore children who are living with their parents but are older than 18. Thirdly, fewer women younger than 18 are mothers, and women younger than age 35 are very unlikely to have children over the age of 18.

I also exclude households who have moved into their current address less than 12 month ago because some of the labor market outcome variables reported in the ACS refers to the year

[9] The Integrated Public Use Microdata Series provides a more user-friendly version of the ACS data. The ACS is conduct by the US Census bureau and is also available on the Census website.

before the time of survey. The dichotomy choice of being a renter or a homeowner usually takes place when the household moves to a new address. For those households that moved into their current addresses fewer than 12 months ago, it is difficult to determine whether the decision on homeownership is before or after the change of their labor market status. This ambiguity may cause unexpected problems.

I further screen out observations in the following categories: male household heads that are not in labor force, self-employed, living in farms or group quarters, currently serving in military, and using the house for commercial purpose. All households included in the sample are residing in the Metropolitan Statistical Areas (MSA), where labor market and housing market are relatively dense and competitive. The final full sample is comprised of 84,805 observations of white male household heads.

Table 1 reports the summary statistics of key variables of the demographic and socio-economic attributes of male household heads. The variable *Homeownership* is a dummy variable with value 1 if the household owns its current home, and zero otherwise. The variable *Employment* is also a dummy. It takes value 1 if the male household head is employed and 0 otherwise. The variable *Hours worked* measures the male household head's usual hours worked per week in the past 12 months before the time of survey. Other variables are used as controls in regression models, they are age and age squared, whether wife is at work, whether the male household head has the U.S. citizenship, and a series of categorical dummy variables indicating the male household head's level of education attainments. The continuous variable *FHFA Index change* is included in all regressions to control for price change in the local housing market. This variable is MSA-specific and is matched with other household-level variables using households' MSA codes. It measures the percentage change of the FHFA index from 2007 to 2010 for each

MSA. The FHFA Index is a quality-adjusted repeat transactions housing price index produced by the Federal Housing Finance Agency at the MSA level.

To further control individual male household head's labor market attributes, all estimation models in this paper include series of dummies indicating their associated two-digit occupation categories and two-digit industry categories. There are 22 occupation effects and 24 industry effects in total.

Table 2 shows the frequencies of different family sizes of the single family households in the sample. Family size is measured by a series of categorical dummies indicating the number of children in the households. The family size categories (from *no child* to *six or more children*) are mutually exclusive and collectively exhaustive. The first two columns show the frequencies and percent of different family sizes for the full sample. The second pair of columns and the third pair of columns report the frequencies and percent of number of children respectively for sample that only includes homeowners and sample that only includes renters.

4. Relationship between homeownership and family size

4.1 A conceptual model

How does homeownership relate to family size? In this section, I explore this relationship based on a conceptual model in which homeownership is jointly determined by the household's consumption demand and investment demand for housing stocks. Both demands are functions of number of children in the household. This model is a version of the theoretical model proposed by Henderson and Ioannides (1983). They assume that housing stock is homogeneous, and can be owned for both consumption and investment purposes, or for investment purpose only. A household has consumption demand for housing stock derived from housing services it wish to consume, and has an investment demand for housing stock based on its portfolio motives. In the absence of tax distortions, borrowing constraints, and transaction costs, families tend to live in owner-occupied housing if their investment demand for housing exceeds their consumption demand for housing (Henderson and Ioannides, 1983).

The conceptual model assumes both the consumption demand and investment demand for housing stock are associated with the family size (i.e., the number of children) of the household. For the consumption demand, it is reasonable to think that families of larger sizes tend to consume more housing services. In a single family household, if there are more children, the household needs more housing space to accommodate the entire family. For the investment demand for housing stock, I assume it is positively and nonlinearly associated with the number of children. There are a couple of reasons. Firstly, households that have more children are more likely to assert control on their homes, such as to be able to freely improve or modify housing space and structures to accommodate the needs of children. They thus have a higher investment demand for housing and are more likely to own their housing property. Secondly, as the number of children becomes larger, raising children will cost a larger part of the family's financial resources. This also imposes a constraint on the household's investment demand for housing, and reduces its likelihood to become homeowner.

Figure 1 illustrates the conceptual model, in which both consumption demand and investment demand for housing are functions of number of children. The horizontal axis indicates the number of children in the household. An increase in number of children increases both the investment and consumption demands of the household. The vertical axis indicates housing stock. In the region *Rent1*, where the household only has a few number of children, its associated consumption demand exceeds the investment demand. It indicates that the amount of housing stock the household wish to consume is higher than the housing stock it wants to assert control for, and the household is better off renting than owning. In the region *Own*, where there are a moderate number of children, it is advantageous for the household to owner-occupy the home since the investment demand exceeds the consumption demand. In this case, being a homeowner provides both the amount of housing services and control needed to accommodate the family. Finally, in the region *Rent2* where the household has a large number of children, its investment demand is impeded by the increasing cost of raising children. This makes the investment demand lower than the consumption demand, and the household is better off to rent rather and owner-occupy its home.

Based on the conceptual model, I expect that households have a moderate number of children (i.e., two or three) are more likely to be homeowners, comparing with households with fewer or more children. This would suggest that the effects of family size on homeownership are highly nonlinear: it first increases homeownership for households with relative small family size, and decrease homeownership for households with very large family size.

4.2 Estimations and results

I use probit models to estimate the relationship between homeownership and family size. The data used for the estimation is the household-level ACS data described in section 3.

The base specification is

$$homeownership_i = \alpha + \sum_{kid=1}^{6} \beta D_{kid,i} + \gamma X_i + u_i \qquad (1)$$

Where the subscript "i" indicates individual male household heads. Recall the dependent variable *homeownership* equals to 1 if the household is a homeowner and 0 otherwise. *X* is a vector of controls for demographic and socio-economic attributes, and γ is the corresponding coefficients. D_{kid} is a series of categorical family size dummies indicating the number of children in the household, and β is the corresponding coefficients.

It is important to note that the omitted category of family size dummies is *no child*, in this case the reference group is households with no child. The family size dummies included in the equation are *one child*, *two children*, *three children*, *four children*, *five children*, and *six or more children*. The estimated coefficients on these family size dummies indicate the relative likelihood of being homeowners comparing to that of households with no child.

Column (1) in Table 3 reports results based on the base specification. The estimated coefficients on family size dummies suggest that households with one to five children are significantly more likely to be homeowners comparing to households with no child. The associated average marginal effects are reported in brackets. By inspecting marginal effects across family size dummies, it is clear that the effects of children on homeownership is highly nonlinear and nonmontonic. Households with *one child* are 8.6% more likely to be homeowners relative to those who have no child. The estimated relative likelihood are 11% for both *two children* and three *children*, and then decreases to 9.7% for *four children* and 5.6% for *five children.* It further drops to 2.5% for *six or more children* and becomes statistically insignificant. Figure 2A plots the average marginal effects of different family sizes, using *no child* as the reference group. The plot indicates that the positive impacts of family size on homeownership are highly nonlinear and nonmonotonic.

As an alternative to the base specification, I estimate a probit model where the omitted category of family size is *two children*; that is

$$homeownership_i = \alpha + \sum_{kid=0}^{1} \beta D_{kid,i} + \sum_{kid=3}^{6} \beta D_{kid,i} + \gamma X_i + u_i \quad (2)$$

where households with two children is the reference group for all other households of different family sizes, including *no child*, *one child,* and from *three children* to *six or more children*. In this specification, the estimated coefficients on family size dummies imply the relative likelihood of being homeowners comparing to that of households with two children.

Column (2) in Table 3 reports the estimation results based on equation (2). The negative and significant coefficients on all specified family size dummies indicate that the corresponding households are less likely to be homeowners comparing to households with two children. Average marginal effects of each family size dummies are reported in brackets, and are plotted in Figure 2B. Note the relative patterns between each family sizes and the nonlinearity and nonmonotonicity are consistent with the plot illustrated in Figure 2A. Households with two children, as the reference group, have the highest likelihood to be homeowners comparing to others. More importantly, the results show that children beyond the second child are negatively and significantly associated with the likelihood of homeownership.

In addition, I experiment a linear specification to test the effects of family size on homeownership for households with more than two children. I include a continuous variable $N_{kid \geq 2}$, which indicating the number of children if the household has more than two kids, and two dummy variables for households with *no child* and *one child*; that is

$$homeownership_i = \alpha + \sum_{kid=0}^{1} \beta D_{kid,i} + \theta N_{kid \geq 2,i} + \gamma X_i + u_i \quad (3)$$

Column (3) of Table 3 reports the results based on the model equation (3). The estimated coefficient on variable $N_{kid \geq 2}$ is negative and statistically significant (at 1% level). The average marginal effect (reported in brackets) indicates that by adding one more child in the households who have already had two children, the likelihood of being homeowners decrease by 1.3%. This result is consistent with the findings from previous nonlinear specifications, and confirms the negative and significant incremental effects of children beyond the second child.

To summarize, the above findings based on different specifications are congruent with the conceptual model, in which the likelihood of a household being homeowner is associated with the numbers of children in a nonlinear and nonmonotonic relationship. Moreover, the likelihood of homeownership is negatively and significantly associated with the number of children for households that have more than two children.

5. The effect of homeownership on men's labor supply

Instrumental variable (IV) is a standard approach to address the problem of endogeneity. However, adequate instrument for homeownership is often difficult to find. A valid instrument for homeownership should be strongly correlated with the propensity of household's housing tenure status, but uncorrelated with the male household head's labor market decisions. There are several instruments that potentially fit these criteria.

5.1 Instrumental variables

A possible aggregate-level instrument is the state marginal tax rate. In the U.S. many states have tax codes that allow deduction of the mortgage interest from individual's income tax payments, while some other states do not permit such deductibility. If the mortgage interest is deductible in a state, individuals in this state are likely to have higher propensity to own homes rather than renting. It is also plausible to argue that the marginal tax rate allowing deduction of the mortgage interest has no effect on individual's labor supply. However, this instrument has some potential limitations. Firstly, the state marginal tax rate is an aggregate-level measure, which may not provide sufficient variations in estimations using cross sectional individual level data. Secondly, it is valid to argue that the state tax policy is an important factor to influence individual's decision in choosing where to live and to work, and tax policies could be potentially interdependent with individual's labor market decisions.

For individual level instruments of homeownership, one example is the same-sex instrument used by Coulson and Fisher (2008) to test homeownership's impact on married men's unemployment and income. The same-sex instrument is originally used by Angrist and Evans (1998) to solve the endogenetiy between fertility and women's labor supply. They use the gender difference of the two first-born children as an instrument for the birth of a third child. This approach is reasonable under the assumption that parents prefer a third child in a different gender if their first two kids are the same sex. Because the gender composition of the first two children is virtually random, it is exogenous to parents' labor market decisions. Coulson and Fisher (2008) assume that having a third child increases households' likelihood to be homeowners, and they argue that the same-sex measure is an appropriate instrument for homeownership. Since it predicts the birth of a third child, it is relevant to homeownership. In addition, it is exogenous to men's labor supply.

However, there are several potential limitations in the same-sex instruments. First of all, the same-sex measure only considers the relative effect of the third child comparing to the first two children. Since the effect of children is highly nonlinear and nonmonotonic, the incremental impact of the third child may not be sufficient for the identification purpose. For instance, estimates based on equation (2) show that the impact of having three children on homeownership does not significantly differ from the impact of having two children.[10] Secondly, using the same-sex instrument requires a sample of households who have at least two children. This restriction reduces the household sample size significantly, to the extent that the final sample only represents a small share of all the single-family households in U.S.[11]

To address the endogeneity problem of homeownership, I use a new set of household-level instruments. They are a set of family size dummy variables indicating whether the household has *three children*, *four children*, *five children*, and *sixth or more children*. In the previous sections, I have shown that the incremental effects of children after the second child are highly correlated with the probability of households being homeowners, and they are possibly exogenous to men's labor supply. Therefore, they have potential to be valid instruments for homeownership. In the following part, a series of IV diagnostic tests are performed to test the validity of these instruments.

5.2 The IV recursive bivariate probit model

I use a recursive bivariate probit model to estimate the effect of homeownership on the probability of being employed. The recursive probit model is a standard econometric method to correct for the endogeneity issue (Greene, 1998). Identification of the effect of homeownership relies on the set of family size instruments.

The model contains two equations:

[10] See results reported in column (2) of Table 3. These results are the probit estimates of the effect of children on homeownership.
[11] In my ACS sample of white male household heads (age from 18 to 55, and with wife's age from 18 to 35) of married couple from single-family households, only 51.2% have two or more than two children in their households.

$$homeownership_i = \alpha + \sum_{kid=0}^{1} \beta D_{kid,i} + \sum_{kid=3}^{6} \beta D_{kid,i} + \gamma X_i + u_i \quad (4)$$

$$employment_i = \tau + \varphi homeowneership + \sum_{kid=0}^{1} \theta D_{kid,i} + \delta X_i + e_i \quad (5)$$

Where *i* is the index for individual household observations associated with the male household heads. *X* is a vector of model controls as described earlier. D_{kid} denotes the set of dummy variables indicating whether the household have *kid* number of children. To test the effect of homeownership on man's labor supply, *homeownership* is included as the instrumented variable in the employment equation (5). The error terms of the two equations u_i and e_i are assumed to have joint normality.

The instruments are four family size dummies (denoted by $\sum_{kid=3}^{6} D_{kid,i}$), indicating whether the household has *three children, four children, five children, and six and more children*. They are included in equation (4) and are excluded from equation (5). It is important to note that the reference group for family size dummies in equation (4) is households with *two children*. The reference group in equation (5) is households with *two and more children*.

The two equations included in the recursive bivariate probit model can be estimated simultaneously using the maximum likelihood method. If endogeneity is present between *homeownership* and *employment*, the error term u_i of the homeownership equation (4) will be significantly correlated with the error term e_i in the employment equation (5). The simultaneous estimation of the model can ensure that this correlation is dealt with, as the correlation matrix of error terms is estimated. Moreover, the dependent variables of equation (4) and equation (5) are both binary outcome variables. The bivariate probit model ensures that the nonlinearity in both equations is properly considered.[12]

5.3 Estimation results and diagnostic tests

The results are reported in Table 4[13]. In column (1), I first show estimates of a conventional probit model, where *homeownership* appears as explanatory variable in the employment equation without considering its potential endogeneity. The estimated coefficient on *homeownership* is 0.335 and is statistically significant, indicating a large and strong correlation between homeownership and men's labor supply. The average marginal effect suggests that male household heads who own their homes are 2.6% more likely to be employed relative to renters.

The estimates of the IV recursive bivariate probit model are reported in column (2) and column (3). The estimates of the first equation (equation 4) are shown in column (2). The estimated coefficients on the instrumental variables are all statistically significant at 1% level. Moreover, the F-statistic of the joint significance test of all instruments is 18.84, with p-value of 0.002. These results imply that the instruments have a strong relevance to the instrumented variable *homeownership*.

Next I carry out a test to examine whether *homeownership* is endogenous to *employment*. Due to the nonlinear nature of the bivariate probit model, the standard Hausman test, which is based on the linear property of the OLS and IV 2SLS estimations, is not appropriate here for the

[12] An alternative to bivariate probit estimation is to use the conventional linear IV estimators. See Chiburis, Das and Lokshin (2011) for a discussion of the similarity and difference between bivariate and linear probability model for IV estimation. In general, the linear results are similar to the ones based on bivariate probit model. However, because the bivariate probit is estimated using maximum likelihood, the estimations can provide better efficiency. In this study the bivariate probit model is preferred over the linear probit model.

[13] All regressions reported in Table 4 contain additional socio-demographic controls for the household characteristics and attributes of the male household head. To conserve space, these controls are not reported in the table. A footnote in the table provides the list of these controls.

testing of endogeneity. Drawing on the work by Knapp and Seaks (1998), I use an alternative approach to test whether the correlation between the residuals of equation (4) and equation (5) is significantly different from zero.[14] A significant non-zero correlation coefficient (denoted as *rho*) rejects the null hypothesis that there is no correlation between *homeownership* and unobserved factors affecting employment. This is equivalent to rejecting the null hypothesis that the *homeownership* is exogenous to employment. In column (3), the Wald statistics (with p-value of 0.063) rejects that the *rho* is equals to zero, which indicates that *homeownership* is indeed endogenous. The result of the Hansen's J-test for over-identification[15] is also reported in column (3). The p-value is 0.56, suggesting that the instruments are likely to be valid.

The key result in column (3) of Table 4 is the estimated coefficients on *homeownership*, the instrumented variable. The positive and significant (at 5% level) coefficient suggests that homeownership has a positive impact on men's labor supply. The average marginal effect suggests that being a homeowner increase the employment probability of male household heads by 1.2%. Comparing results between the IV marginal effect (1.2%) and the probit marginal effect (2.6%), the difference suggests that there is a substantial upward endogeneity bias in the conventional probit estimation. The economic magnitude of the probit estimate almost doubles that of the IV estimate.

The IV result is consistent with the theory that homeownership tends to increase labor market participation, for reasons such as mortgage debt commitment and bearing of additional ownership costs. It offers evidence to support the previous findings that homeownership has a

---

[14] Knapp and Seaks (1998) show that a likelihood-ratio test of whether the correlation coefficient of the residuals of the two equations in a bivariate probit model is equal to zero can be used as a Hausman endogeneity test.
[15] Since there is no proper method to directly perform the over-identification test in the bivariate probit model, instead I perform the test using the IV linear probability model.

positive effect on labor supply. However, the IV results do not support the hypothesis proposed by Osward (1997), in which homeownership is associated with inferior labor market outcomes.

5.4 Results based on samples of different age groups

Age is an important factor that affects both homeownership and men's labor supply. Individual household's transition from renter to homeowner is closely related to the age of the household head. It is also well documented that men's labor supply is affected by their age. Thus I expect that for male household heads in different stages of life, their labor supply may respond differently to the impact of homeownership.

In this section, I consider how the effect of homeownership on men's labor supply differs between households of different age groups. I stratify the full sample of male household heads into two age groups: one group is for those with age between 18 and 33, and the other group is for those with age between 34 and 55. The cut off line 33 is the average age in the full sample, which generates two age groups with equal sample size. For each age group, I repeat the IV recursive bivariate probit estimation using the same set of instrumental variables.

Table 5 reports the IV results for different age groups of male household heads. Note that only the estimated coefficients on the instrumented variable *homeownership* are reported, the coefficients on other control variables are not reported in the table. Column (1) reports the results based on the group of household heads aged 18-33. The coefficient is positive and significant, indicating a positive effect of homeownership on younger men's labor supply. The average marginal effect suggests that younger male household heads who own homes are 1.3% more likely to be employed relative to renters. However, for male household heads aged 34-55, the IV estimate in column (2) indicates that homeownership has no significant impact on their labor supply. Comparing results from age group 18-33 and age group 34-55, it reveals that homeownership's impact on men's labor supply tends to be concentrated on younger male household heads. For male household heads that have reached their mid-30's, homeownership seems to have little impact on their labor supply.

The usual set of IV diagnostic tests is performed for each age group, and reported in Table 5. They include the F test on the relevance of the instruments, the test on the endogeneity of homeownership, and the over-identification test. The results of these tests suggest that the instruments and the bivariate probit model are appropriate for the IV estimations.

6. Relationship between men's hours worked and family size

In this section, I investigate the relationship between family size and men's hours worked. I estimate a reduce form labor supply function using usual hours worked as the dependent variable. That is

$$log(hours_i) = \alpha + \sum_{kid=0}^{1} \beta D_{kid,i} + \sum_{kid=3}^{6} \beta D_{kid,i} + \gamma X_i + u_i \qquad (6)$$

where the dependent variable is the logged usual hours worked per week of male household heads. The reference group (the omitted category) of family size dummies is households with two children. The data used for the estimation is the sample of married white male household heads described in section 3. Note that male household heads with zero hours worked are not used in the estimation (582 observations dropped[16]).

[16] The size of the male household heads sample used in the hours worked estimation is 84,218.

Column (1) of Table 6 reports the OLS estimates on family size dummies, based on equation (6)[17]. Figure 3 plots these estimates to illustrate the nonlinearity of the effect of family size on hours worked. The result suggests that the incremental effect of the third and additional children on men's hours worked is negative and significant. For households with two or more children, having additional kids tends to reduce men's hours worked so that they could spend more time caring for children at home. Since the third and additional children are significantly associated with men's hours worked, excluding them from equation (6) is likely to cause heterogeneity bias. Therefore, the third and additional children may not be appropriate instruments to estimate the effect of homeownership on men's hours worked.

With the IV approach out of the question, I have to rely on OLS to look for the influence of homeownership on men's hours worked. Caution is required in interpreting OLS results because the endogeneity problem is not address in OLS estimations. Column (2) of Table 6 reports OLS estimates of an hours worked regression which includes *homeownership* as a regressor. The estimated coefficient suggests that homeowners work 4.8% more hours per week relative to renters. Although this estimate is likely to be biased, it gives a rough sense of the direction and upper bond of the effect of homeownership. Moreover, in Table 6 column (3) and (4) report OLS results based on two sub-samples of different age. One group is men of age 18-33, and the other is men of age 34-55. Comparison between the two groups suggests that homeownership tend to have a larger influence among younger male household heads (with 5.2% increase in hours worked) relative to older ones (with 4.0% increase in hours worked). This relative pattern is consistent with earlier results of homeownership's effects on the employment

[17] Estimated coefficients on socio-economic controls, such as age and education levels, are not reported in the table.

of younger versus older men.[18] Perhaps for men younger than 33, being a homeowner (or want to be a homeowner) provides more motivation and incentive to work harder, relative to men in older age.

7. Conclusion

In this paper, I show that the relationship between family size and homeownership is nonlinear and nonmonotonic for single family households in the U.S. For households with two or less children, homeownership is positively correlated with family size. But for households that have more than two children, homeownership is negatively correlated with family size.

Using the presence of children after the second child as instruments for homeownership, I adopt an IV strategy to solve the endogeneity problem between homeownership and men's labor supply. The IV results suggest that households who are homeowners are 1.2% more likely to be employed than renters. The magnitude of the IV estimate is half of the magnitude of the conventional probit estimate. Moreover, I estimate the same IV model using samples stratified by age group. The results indicate that homeownership has positive and significant effect on men's labor supply for younger male household heads, but not for older men.

My findings have contributions as well as limitations. In this paper, I use a new set of instruments to correct the endogeneity between homeownership and men's labor supply. Previous empirical literature has limited success in solving this endogeneity problem, but the instruments used in this paper seem to produce valid and sensible results. This paper also has several limitations. The estimations in this paper use cross-sectional data, which may have

[18] See Table 6 for IV estimates of the effect of homeownership on men's employment, based on samples stratified by age groups.

potential heterogeneity problem caused by unobserved factors. Future research using better dataset may shed further light on this issue. In addition, although this paper provides evidence of the positive effect of homeownership on men's labor supply, the specific mechanism of this effect is still unclear and requires further research.

References

Angrist, J. D., and W. N. Evans, (1998) "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size", *American Economic Review*, 88(3), 450-77.

Barrios, J. A. and J. E. Rodríguez, (2004) "User cost changes, unemployment and homeownership: evidence from Spain", *Urban Studies*, 41, pp. 563-578.

Bottazzi, R., (2007) "Labour market participation and mortgage-related borrowing constraints", Institute for Fiscal Studies, *mimeo*

Bottazzi, R., H. Low, and M. Wakefield, (2007) "Why do home owners work longer hours?" IFS Working Papers W07/10, Institute for Fiscal Studies.

Chan, S., (2001) "Spatial lock-in: do falling house prices constrain residential mobility?" *Journal of Urban Economics,* 49 (3), 567-586.

Chiburis, R. C., J. Das, and M. Lokshin, (2011) "A practical comparison of the bivariate probit and linear IV estimators", Policy Research Working Paper Series, The World Bank.

Coulson, N. E. and L.M. Fisher, (2002) "Labor market outcomes of homeownership", *Housing Studies,* 17, 35-50 (January).

Coulson, N. E. and L.M. Fisher, (2008) "Housing tenure and labor market impacts: The search goes on", *Journal of Urban Economics,* 65, 252-264.

Courgeau, D., and E. Lelievre, (1992) "Interrelations between first home-ownership, constitution of the family, and professional occupation in France", In J. Trussell, R. Hankinson, & J. Tilton (Eds.), *Demographic Applications of Event History Analysis* (pp.120–140). Oxford: Clarendon Press.

Dipasquale, D., and E. Glaeser, (1999) "Incentives and Social Capital: Are Homeowners Better Citizens?" *Journal of Urban Economics*, 45(2), 354–384.

Dohmen, T. J., (2005) "Housing, mobility and unemployment", *Regional Science and Urban Economics,* 35, 305-325.

Del Boca, D. and A. Lusardi, (2003) "Credit Market Constraints and Labor Market Decisions", *Labor Economics*, 10: 681-703.

Engelhardt, G.V., (2003) "Nominal loss aversion, housing equity constraints, and household mobility: evidence from the United States", *Journal of Urban Economics*, 53 (1), 171-195.

Feijten, P., and C. H. Mulder, (2002) "The timing of household events and housing events in the Netherlands: A longitudinal perspective", *Housing Studies*, 17(5), 773–792.

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, P. Hendershott, G. Wood, (2003) “Homeownership and unemployment: The roles of leverage and public housing”, Working paper 10021, NBER.

Fortin, N. M., (1995) “Allocation inexibilities, female labor supply, and housing assets accumulation: Are women working to pay the mortgage?” *Journal of Labor Economics*, 13 (3), 524-57.

Genesove, D. and C. Mayer, (1997) “Equity and time to sale in the real estate market”, *American Economic Review*, 87 (3), 255-269.

Genesove, D., C. Mayer, (2001) “Loss-aversion and seller behavior: evidence from the housing market”, *Quarterly Journal of Economics,* 116 (4), 1233-1260.

Green, R. and P. Hendershott, (2001) “Home ownership and unemployment in the U.S.”, *Urban Studies,* 38, 1509-1520.

Greene, W.H., (1998) “Gender Economics Courses in Liberal Arts Colleges: Further Results”, *Journal of Economic Education*, 29(4), pp. 291-300.

Henderson, J. V., and Y. M. Ioannides, (1983) “A Model of Housing Tenure Choice”, *American Economic Review*, 73 (March), 98±113.

Knapp LG, and TG. Seakes, (1998) “A Hausman Test for a Dummy Variable in Probit”, *Applied Economics Letters,* 5, 321–23.

Kulu, H., and A. Vikat, (2007) “Fertility differences by housing type: The effect of housing conditions or selective moves?” *Demographic Research*, 20, 775–802.

Lundberg, S., E. Rose, (2002) “The effects of sons and daughters on men’s labor supply and wages”, *Review of Economics and Statistics,* 84, 251-268.

Mulder, C. H., and M. Wagner, (2001) “The connections between family formation and first-time homeownership in the context of west Germany and the Netherlands”, *European Journal of Population*, 17, 137–164.

Munch, J., M. Rosholm, and M. Svarer, (2006) “Are homeowners really more unemployed?” *Economic Journal*, 116, 99-1013.

Munch, J., M. Rosholm, and M. Svarer, (2008) “Home ownership, job duration, and wages”, *Journal of Urban Economics,* 63, 130-145.

Murphy, M. J., and O. Sullivan, (1985) “Housing tenure and family formation in contemporary Britain”, *European Sociological Review*, 1(3), 230–243.

Oswald, A., (1997) “Thoughts on NAIRU” (Correspondence), *Journal of Economic Perspectives,* 11, 227-228.

Oswald, A., (1999) “The housing market and Europe’s unemployment: A nontechnical Paper”, Working paper, University of Warwick.

Partridge, M., and D. Rickman, (1997) “State unemployment differentials: Equilibrium factors vs. differential employment growth”, *Growth and Change,* 28, 360 – 379.

Pehkonen, J., (1999) “Unemployment and homeownership”, *Applied Economics Letters,* 6, 263-265.

Quigley, J. M., (1987) “Interest rate variations, mortgage prepayments and household mobility”, *Review of Economics and Statistics,* 69, 636-643.

Rohe, W. M., and L. S. Stewart, (1996) “Homeownership and neighborhood stability”, *Housing Policy Debate,* 7(1), 37.81.

Stein, J, (1995) “Prices and trading volume in the housing market: a model with down-payment effects”, *Quarterly Journal of Economics,* 110, 379-406.

TABLE 1

Descriptive Statistics[a]

Dummy variables	Frequency	Percent
Homeownership	66,909	78.9
Employment	81,977	96.7
Spouse is at work	57,457	67.8
Less than highschool	6,250	7.4
Highschool GED or diploma	32,107	37.9
Associate degree	7,249	8.6
Bachelor degree	26,130	30.8
Master degree	9,214	10.9
Professional degree	2,611	3.1
PhD degree	1,244	1.5
U.S. citizenship	76,825	90.6
Continuous variables	Mean	Std. Dev.
Hours Worked	44.5	10.0
Age	33.8	5.4
FHFA Index change	-0.1	0.1

a. There are a total of 84,805 white male household heads included in the sample. All included households live in 297 MSAs. Sample drawn from the ACS 2010 three year 3% file.

TABLE 2

Frequency Distribution: Number of Children in the Household[a]

	Full Sample		Homeowners		Renters	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Number of children in the household						
No children	20,874	24.61	15,230	22.76	5,644	31.54
One child	20,590	24.28	16,519	24.69	4,071	22.75
Two children	27,499	32.43	22,797	34.07	4,702	26.27
Three children	11,569	13.64	9,214	13.77	2,355	13.16
Four children	3,239	3.82	2,432	3.63	807	4.51
Five children	705	0.83	492	0.74	213	1.19
Six or more children	329	0.39	225	0.34	104	0.58

a. The sample includes a total of 84,805 households, with white male household head between the ages of 18 and 55.

TABLE 3
The Effect of Family Size on Homeownership
Dependent Variable: Homeownership (1=owner, 0=renter)[ab]
Probit Model Estimates, with Standard Errors in Parentheses and Average Marginal Effects in Brackets[c]

	(1)	(2)	(3)
No child	-	-0.460***	-0.471***
	-	(0.025)	(0.027)
	-	[-0.116]	[-0.119]
One child	0.328***	-0.131***	-0.143***
	(0.017)	(0.018)	(0.020)
	[0.086]	[-0.029]	[-0.032]
Two children	0.460***	-	-
	(0.025)	-	-
	[0.116]	-	-
Three children	0.451***	-0.008	-
	(0.031)	(0.018)	-
	[0.114]	[-0.001]	-
Four children	0.376***	-0.084***	-
	(0.044)	(0.033)	-
	[0.097]	[-0.018]	-
Five children	0.207***	-0.253***	-
	(0.063)	(0.069)	-
	[0.056]	[-0.059]	-
Six or more children	0.088	-0.372***	-
	(0.110)	(0.122)	-
	[0.025]	[-0.091]	-
Number of children (2~6, continuous)	-	-	-0.056***
	-	-	(0.015)
	-	-	[-0.013]
Spouse is at work	0.333***	0.333***	0.332***
	(0.014)	(0.014)	(0.014)
Age	0.236***	0.236***	0.237***
	(0.009)	(0.009)	(0.009)
Age squared	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)
Highschool GED	0.319***	0.319***	0.319***
	(0.033)	(0.033)	(0.033)
Associate degree	0.569***	0.569***	0.568***
	(0.037)	(0.037)	(0.037)
Bachelor degree	0.690***	0.690***	0.689***
	(0.036)	(0.036)	(0.036)
Master degree	0.664***	0.664***	0.663***
	(0.042)	(0.042)	(0.042)
Professional degree	0.553***	0.553***	0.551***
	(0.068)	(0.068)	(0.068)
PHD degree	0.376***	0.376***	0.374***
	(0.070)	(0.070)	(0.070)
U.S. citizenship	0.792***	0.792***	0.790***
	(0.035)	(0.035)	(0.034)
FHFA index change	0.932***	0.932***	0.933***
	(0.242)	(0.242)	(0.242)
F-test on all variables of more than two children	-	18.84	13.99
p-value	-	{0.001}	{0.001}
No. of Occupation effects	22	22	22
No. of Industry effects	25	25	25
No. of observations	84,805	84,805	84,805
Pseudo R squared	0.16	0.16	0.16

a. Estimations are based on 2010 ACS three years sample, restricted to married white male household head between age 18 and 55, whose wife is between age 18 to 35, and are living in households in MSAs.
b. *** 1% statistical significance, ** 5% statistical significance, * 10% statistical significance.
c. In parentheses "()" reports the Robust Standard Errors adjusted for 276 MSA clusters. In Brackets "[]" reports the average marginal treatment effects for the sample population. In curly brackets "{}" reports p-value.

TABLE 4
The Effect of Homeownership on Men's Labor Supply[ab]
IV Bivariate Probit Estimates, with Standard Errors in Parentheses and Marginal Effects in Brackets[cd]

	Probit Model	IV Bivariate Probit Model	
		1st Equation	2nd Equation
Dependent Variable:	Employment	Homeownership	Employment
	(1)	(2)	(3)
Homeownership	0.330***	-	0.166**
	(0.020)	-	(0.081)
	[0.026]	-	[0.012]
No children	-0.045*	-0.460***	-0.064**
	(0.023)	(0.025)	(0.025)
	[-0.003]	[-0.116]	[-0.004]
One children	-0.050**	-0.131***	-0.054**
	(0.023)	(0.018)	(0.023)
	[-0.003]	[-0.029]	[-0.003]
Two or more children (reference group)	-	-	-
Instruments excluded from the 2nd equation			
Three children	-	-0.008	-
	-	(0.018)	-
	-	[-0.001]	
Four children	-	-0.084***	-
	-	(0.033)	-
	-	[-0.018]	
Five children	-	-0.253***	-
	-	(0.069)	-
	-	[-0.059]	
Six or more children	-	-0.372***	-
	-	(0.122)	
	-	[-0.091]	-
F-statistic on excluded instruments	-	18.84	-
p-value	-	{0.002}	-
Endogeneity test (Wald test of rho=0)	-	-	3.38
p-value	-	-	{0.063}
Hensen-J test for over-identification	-	-	2.82
p-value	-	-	{0.424}
No. of observations	84,805	84,805	84,805

a. Estimations based on 2010 ACS three years sample, restricted to married white male between age 18 and 55, whose wife is between age 18 to 35, and resides in MSAs.
b. Additional controls included in the regressions but not reported in the table: Spouse is at work, Age, Age squared, Highschool GED, Associate degree, Bachelor degree, Master degree, Professional degree, PHD degree, U.S. citizenship, FHFA Index change, 22 Occupation effects, and 25 Industry effects.
c. *** indicates 1% significance, ** indicates 5% significance, * indicates 10% significance.
d. In parentheses "()" reports the Robust Standard Errors adjusted for 276 MSA clusters. In Brackets "[]" reports the Average Marginal Effects is the average marginal treatment effects for the sample population. In curly brackets '{}" reports p-value.

TABLE 5

The Effect of Homeownership on Men's Labor Supply: Age Group 18-33 vs. Age Group 34-55

Dependent Variable: Employment Status (1=Employed, 0=Unemployed)[ab]

IV Bivariate Probit Estimates, with Standard Errors in Parentheses and Marginal Effects in Brackets[cd]

	Age 18-33 (1)	Age 34-55 (2)
Homeownership	0.192** (0.100) [0.013]	-0.057 (0.122) [-0.004]
F-statistic on excluded instruments	15.02	20.28
p-value	{0.004}	{0.001}
Endogeneity test (Wald test of rho=0)	2.961	7.478
p-value	{0.085}	{0.006}
Hensen-J test for over-identification	3.096	0.424
p-value	{0.377}	{0.935}
No. of observations	42,083	42,722

a. Estimations based on the 2010 ACS three years sample of married white male household heads, stratified by age to groups, age 18-33, and age 34-55.

b. Additional controls included in the regressions but not reported in the table: Spouse is at work, Age, Age squared, Highschool GED, Associate degree, Bachelor degree, Master degree, Professional degree, PHD degree, U.S. citizenship, FHFA Index change, 22 Occupation effects, and 25 Industry effects.

c. *** indicates 1% statistical significance, ** indicates 5% statistical significance, * indicates 10% statistical significance.

d. In parentheses "()" reports the Robust Standard Errors adjusted for 276 MSA clusters. In Brackets "[]" reports the Average Marginal Effects is the average marginal treatment effects for the sample population. In curly brackets '{}" reports p-value.

TABLE 6

The Effect of Family Size on Hours Worked of Male Household Heads[ab]

Dependent Variable: (Log of The) Hours Worked Per Week

OLS Estimates, with Standard Errors in Parentheses[cd]

	Full Sample		Age 18-33	Age 34-55
	(1)	(2)	(3)	(4)
Homeownership	-	0.048***	0.052***	0.040***
	-	(0.003)	(0.004)	(0.003)
No child	-0.005**	-0.001	0.003	-0.002
	(0.002)	(0.002)	(0.003)	(0.003)
One child	-0.004**	-0.003	0.004	-0.011**
	(0.002)	(0.002)	(0.003)	(0.002)
Two children (reference group)	-	-	-	-
	-	-	-	-
Three children	-0.002	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.003)	(0.002)
Four children	-0.009**	-0.008**	-0.008	-0.008*
	(0.004)	(0.004)	(0.007)	(0.005)
Five children	-0.024**	-0.021*	-0.032	-0.017*
	(0.011)	(0.011)	(0.024)	(0.009)
Six or more children	-0.049**	-0.045**	-0.027	-0.054**
	(0.023)	(0.022)	(0.033)	(0.023)
No. of observations	84,218	84,218	41,849	42,369
R squared	0.07	0.08	0.07	0.08

a. Estimations based on 2010 ACS three years sample, restricted to married white male between age 18 and 55, whose wife is between age 18 to 35, and resides in MSAs.

b. Additional controls included in the OLS regressions but not reported in the table: Spouse is at work, Age, Age squared, Highschool GED, Associate degree, Bachelor degree, Master degree, Professional degree, PHD degree, U.S. citizenship, FHFA index, 22 Occupation effects, and 25 Industry effects.

c. *** indicates 1% significance, ** indicates 5% significance, * indicates 10% significance.

d. In parentheses "()" reports the Robust Standard Errors adjusted for 276 MSA clusters. In curly brackets "{}" reports p-value.

e. The sample used for hours worked regression excludes male household heads who are not employed. Thus have a slightly smaller sample size.

Figure 1: Relationship between Family Size and Homeownership

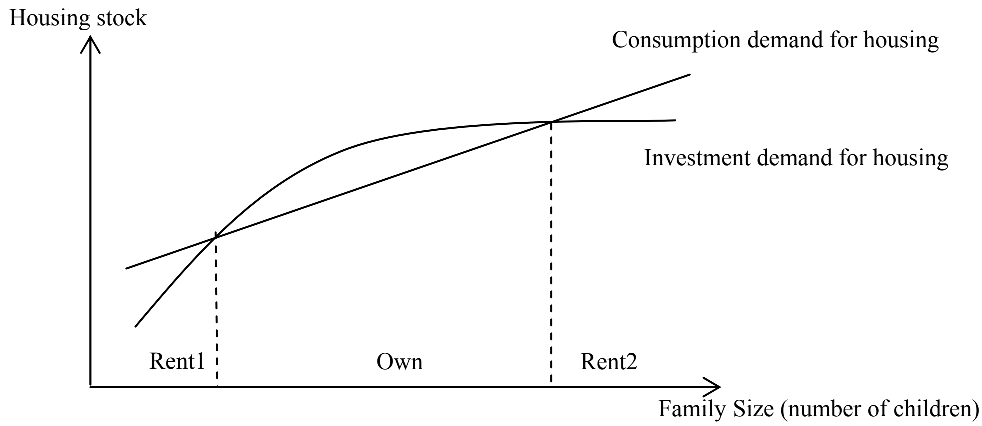


Figure 2A: The Effect of Family Size on Homeownership (Reference Group: No Child)

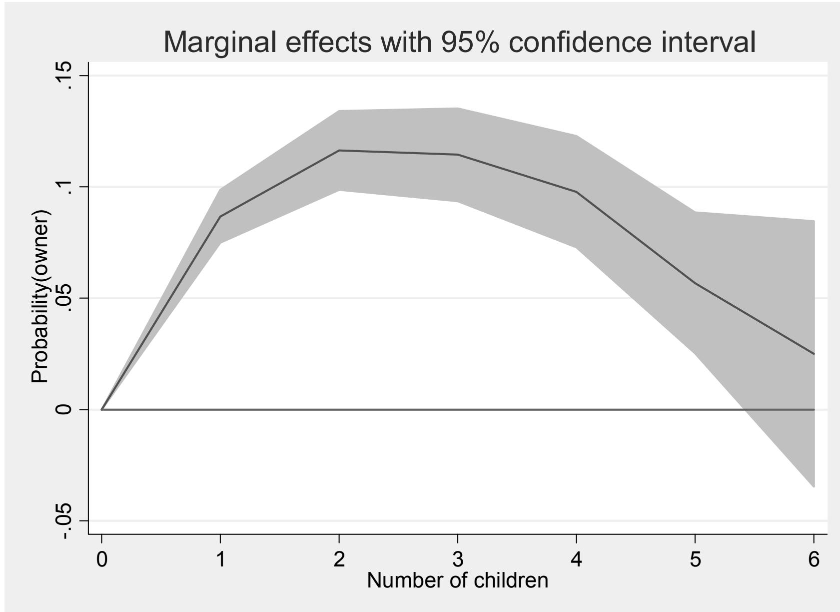


Figure 2B: The Effect of Family Size on Homeownership (Reference Group: Two Children)

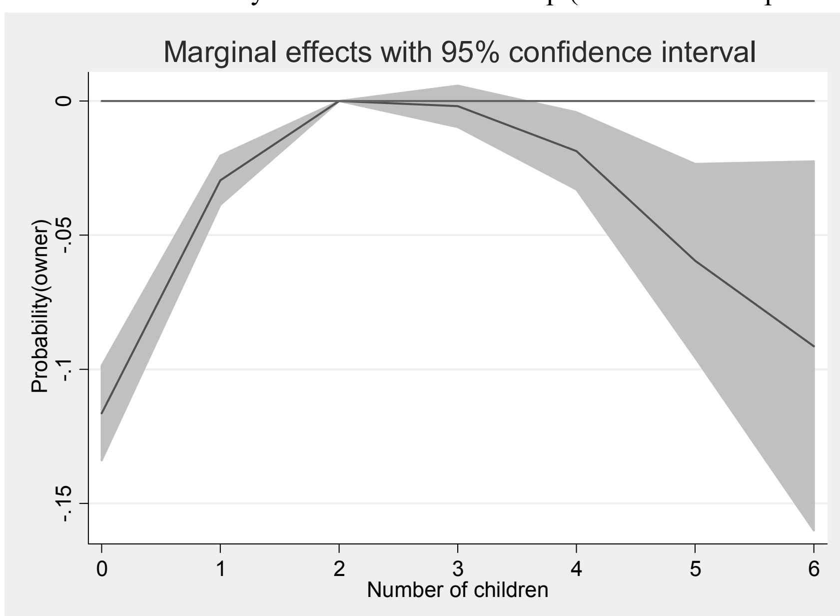
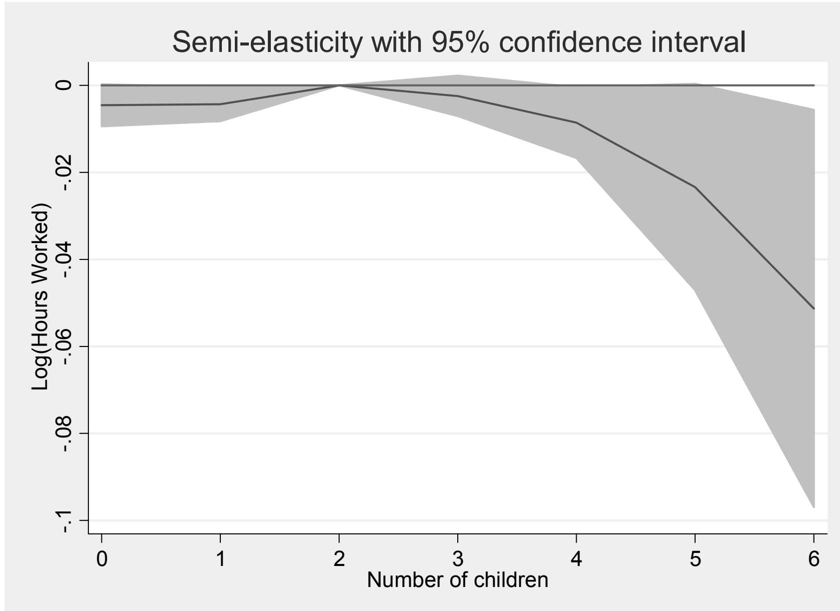


Figure 3: The Effect of Family Size on Men's Hours Worked (Reference Group: Two Children)



VITA

Name of Author: Jie Dai

Place of Birth: Zhejiang, China

Date of Birth: April 18th, 1980

Graduate and Undergraduate Schools Attended:

Syracuse University, Syracuse, New York

Zhejiang University, Hangzhou, China

Degrees Awarded:

Master of Arts in Economics, 2008, Syracuse University

Master of Arts in International Relations, 2008, Syracuse University

Bachelor of Science in Biotechnology, 2002, Zhejiang University

Professional Experience:

Teaching Assistant, Economics Department, Syracuse University, 2008-2013

Graduate Assistant, International Relations Department, Syracuse University, 2006-2008

Associate, Taishan Capital Corporation, Beijing, China, 2002-2006