THREE ESSAYS IN APPLIED ECONOMICS ON

LABOR SUPPLY DECISIONS

By

ZHENGYU CAI

Bachelor of Economics
Southwest Jiaotong University
Chengdu, China
2010

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2016

THREE ESSAYS IN APPLIED ECONOMICS ON

LABOR SUPPLY DECISIONS

Dissertation Approved:

Dan S. Rickman

Dissertation Adviser

John V. Winters

Mehtabul Azam

Jonathan Charles Comer

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Dan S. Rickman for his professional, patient, and painstaking guidance on this dissertation. I also want to thank the committee members for their comments and help. I am very grateful to my family and friends for the support over the past five years, especially to WT for my job seeking, and to M for the sweet and warm companion in the last days of my PhD life.

Name: ZHENGYU CAI

Date of Degree: MAY, 2016

Title of Study: THREE ESSAYS IN APPLIED ECONOMICS ON LABOR SUPPLY DECISIONS

Major Field: ECONOMICS

Abstract: The present study contains three applied economic essays which empirically investigate regional and educational factors influencing labor supply decisions. The first essay uses minimum distance from work PUMA centroid to the US coastlines and estimated industry share in 1930 as instrumental variables, and tests the causal impact of agglomeration on work intensity of the self-employed. The 2SLS results show that only localization has a positive impact on hours worked of the self-employed. Urbanization does not affect hours worked. The causality mostly comes from competition within industries.

The second essay uses 1% IPUMS ACS 2013 data to study self-employment differential between foreign STEM graduates and non-STEM graduates. The empirical results show that the differential is still quite substantial after controlling for several covariates. Furthermore, self-employment differentials across broad major groups and detailed majors are examined. We try to explain the self-employment differentials through the differences in incomes between self-employed and salaried foreign college graduates. Our empirical results show that foreign STEM graduates are less likely to be self-employed, since they could earn significantly more in salaried jobs, but the income advantage disappears when they shift to self-employment. This paper implies that, on one hand, policy makers could consider lowering the immigration barrier for graduates in non-STEM fields with high self-employment rates, or at least reduce the institutional discrimination between STEM and non-STEM graduates. On the other hand, target-based subsidies and/or tax benefits could be offered to the startups co-funded by foreign STEM graduates and those who are educated in fields with high self-employment rates.

The third essay uses instrumental variable estimation to examine two-way causality between human capital externalities and agglomeration. The empirical results show that human capital externality has much more causal impact on agglomeration than the reverse. The relationship between human capital externality and urbanization is much stronger than that with localization for both directions of causality. Then the relationship is used to analyze the causality between human capital spillover and hours worked.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

HOURS WORKED OF THE SELF-EMPLOYED AND AGGLOMERATION

1 Introduction

Agglomeration economies refer to economies of scale that are external to the firm. As an important production factor, does the labor input (supply) relate to such a spillover effect? Rosenthal and Strange (2008) gave this question a positive answer using a salaried worker sample. They find that professional employees work more in localized areas because of the urban rat race effect. However, a higher level mechanism was neglected from their research that employees are organized, managed by the firms (employers), which raises two further research questions regarding whether the same relationship exists for employers? If employers and employees share the same work intensity pattern in clusters, could they be explained by the same mechanism? This paper tries to answer these two questions by using the Integrated Public Use Microdata Series (IPUMS. Ruggles et al., 2010).

Very few studies looked at the association between agglomeration and work intensity either in labor economics or urban economics. Rosenthal and Strange (2008) document the overlooked relationship between agglomeration economies and hours worked. They find that this relationship varies by nature of the work and also by age. For nonprofessionals, the association between hours worked and the extent of localization is negative. They use “work-spreading” to explain this negative relationship. The idea is that when the amount of work is limited, increasing the number (density) of workers will decrease their work intensity. While for professional employees, the association between localization and work intensity is positive. They assert a possible mechanism that under a competitive environment, employees tend to work more in order to signal the authorities (supervisors or bosses) that they are more hardworking/capable than others, then they could have chances to get ahead, which is the so called “urban rat race” effect by Rosenthal and Strange.

However, in an agglomeration context, the endogeneity issue should not be ignored. To solve the endogeneity issue, an instrumental variable (IV) estimation strategy and the use of panel data are two standard methods (Ciccone and Hall, 1996; Glaeser and Maré, 2001; Combes, Duranton, Gobillon, and Roux, 2010). This paper uses an instrumental variable estimation strategy to take care of the endogeneity issue for a cross-sectional sample.

By answering the two questions in the beginning, this paper contributes to the literature in several ways. Firstly, to the first question, this paper shows that the relationship between agglomeration and work intensity for the self-employed is similar as it for employees, which is consistent with Rosenthal and Strange (2008). Levine and Rubinstein (2012) find that the self-employed work more than their salaried counterparts. Thus, if the self-employed who are usually tagged as hard working sort into agglomeration economies, it could drive the positive association between hours worked and agglomeration. After controlling for a set of exhaustive covariates, amenities, industry fixed effects, and MSA fixed effects, the OLS results show that localization is positively correlated with work intensity. Furthermore, the Two Stage Least Squares (2SLS) results confirm the OLS results. These results provide some evidence that clusters could increase regional economic activities, where firms enjoy the positive spillover effects.

Second, to answer the second question, it is impossible to use the same mechanism (urban rat race) to explain the seemingly similar relationship for the self-employed even without considering the endogeneity issue. The reason is simple and straightforward. It is because the self-employed are their own bosses. Therefore, the urban rat race effect makes no sense in the self-employment context. Instead, as this paper will show, the competition within industries and specialization may explain most of the positive relationship.

Third, this paper uses a geographic attribute as one of the instrumental variables for the agglomeration measures. Minimum distance from the work PUMA centroid to the coastline has several advantages as an instrument. First, it is strongly correlated with the agglomeration measure, which excludes the possibility of a weak instrument. Combes et al. (2010) use a geology variable instrument for population density, but it turns out to be a weak instrument. Weak instruments could lead to more severe bias than OLS. Second, minimum distance from the work PUMA centroid to the coastline is less likely to be causally correlated with work intensity. Furthermore, our instrument is also valid controlling for work PUMAs with harbors and ports present for the concern that the instrument affects the dependent variables through ports and harbors.

Moreover, estimated industry share in 1930 is used as the second instrumental variable for the agglomeration measures. To our knowledge, it is rare to simultaneously include urbanization and localization in a single regression since both measures suffer from the endogeneity issue. Thanks to the two valid instrumental variables, we are able to mitigate the endogeneity for two agglomeration measures concurrently. One interesting result is that when localization is solely in the regression, it actually picks up the effect of urbanization. Thus, urbanization and localization may suffer from collinearity. To deal with this issue, we use a method similar to the location quotient, i.e., including the quotient of localization and urbanization instead of the raw localization measure in the regressions.

Fourth, this paper also tests the agglomeration wage effect as a mechanism for the positive relationship between work intensity and agglomeration. We innovatively control for regression-adjusted wage for employees in the regressions for the self-employed to eliminate the mechanism endogeneity. However, controlling for the agglomeration wage effect does not alter our results much. According to a simple labor supply model, it is probably because the income effect balances the substitution effect, so that the agglomeration wage effect does not influence work intensity. This also could be explained that after controlling for demographic characteristics and using IV to take care of the unobserved characteristics, people are neutral in preference.

Lastly, this paper uses Geographic Information System (GIS) data to construct a series of amenity measures at the work PUMA level. Amenity and productivity shocks are capitalized into wage and rent differentials under the spatial equilibrium assumption (Roback, 1982). Hours worked as a labor market outcome is highly likely to be influenced by those shocks. Regression-adjusted wages and rents are usually included to control for the shocks. However, these controls are potentially endogenous (Rosenthal and Strange, 2008; Winters, 2013; Kahn and Lang, 1991). Rosenthal and Strange (2008) use a reduced-form specification to control for wages indirectly. Nonetheless, Kahn and Lang (1991) do not provide very valid support for the reduced form approach. This paper tests the agglomeration wage effect as a possible mechanism for the relationship between work intensity and agglomeration. Controlling for regression-adjusted wages may pick up the agglomeration wage effect rather than amenity and productivity shocks. Therefore, amenities are directly included in our empirical specifications. The development of GIS allows us to construct amenity measures at smaller regional scales.

The next section provides the conceptual framework. Section 3 discusses the empirical framework used by this paper, including model concerns and solutions. Section 4 introduces the data and variable construction. Section 5 provides the empirical results. The last section concludes.

2 Conceptual Framework

The self-employed work for themselves, thus they are their own employers. It is not reasonable for them to work more hours in order to signal themselves to get ahead (getting promotion, getting higher income, etc.). Therefore, the urban rat race effect is less likely to be a convincing driving factor for the self-employed to work more in clusters. Instead, there are several possible ways that could lead the self-employed to work more in agglomeration economies. We will discuss them in this section.

2.1 Urbanization and Localization

Before we proceed to the possible mechanism, it is important to discuss how we define agglomeration economies first. By definition, urbanization economies are agglomeration economies across industries, but within an industry for localization economies. In the literature, there are no perfect measures for urbanization and localization economies. For urbanization, researchers usually use population or employment density as a proxy of urbanization economy. For localization, the industry share or industry-specific employment density are often used. In this paper, we will use population density as the proxy for urbanization, and employment density of a given industry for localization.

2.2 Competition and Specialization

Although the urban rat race cannot be applied, rivals still exist for self-employment. The benefits of localization bring more firms into the clusters, and hence the competition within industries. In order to smooth the benefits (profits), people will naturally increase their work intensity when competition increases. The localization variable mentioned above is a natural candidate to capture such rivalry. On the other hand, specialization leads less competition across industry boundaries, although it enhances productivity and wages. Therefore, urbanization is less likely to increase work intensity, while localization could make people work more.

2.3 Sorting and Simultaneity

Workers who have higher work intensity might sort into work PUMA and/or industries based on unobservable characteristics of the individual and/or area. These unobservable characteristics could drive the work intensity pattern of the self-employed in agglomeration economies. For instance, people who have a taste for longer hours worked may sort into the self-employed for full compensation of their efforts. Also, if agglomeration raises productivity, the self-employed, who tend to work longer hours, would sort into a denser area for greater compensation. Besides, hardworking self-employed individuals could also be attracted by large city amenities.

Moreover, a causal relationship running from hours worked to agglomeration may exist as well. Longer hours worked could bring more competition, higher human capital level, more mature markets, etc., which cause the formation of agglomeration economies. If simultaneity exists, the descriptive relationship between hours worked and agglomeration could be biased.

In order to solve the endogeneity issue, a classic method is using an instrumental variable estimation strategy. This paper will rely on two novel instrumental variables to take care of the sorting and simultaneity issue.

2.4 Agglomeration Wage Effect

A simple labor supply model presents another possible channel. In a world with homogenous preferences, a worker chooses leisure/work and a composite good. The worker only earns an income from work, and the income rate is exogenous. It is easy to know that, if the substitution effect (SE) dominates the income effect (IE), the higher the income, the higher the price of leisure. Therefore, less leisure will be consumed, and more labor will be supplied.

It is worth noticing that if the income effect (IE) dominates the substitution effects (SE), the result will flip the sign, which will generate a negative relationship between incomes and hours worked. Table 1.1 summarizes all the situations.

Table 1.1: Hours Worked and Different Effects

	SE dominates	IE dominates
Income Increases	Hours worked↑	Hours worked↓
Income Decreases	Hours worked↓	Hours worked↑

A lot of studies substantiate that agglomeration increases productivity and wages (Ciccone, 2002; Rosenthal and Strange, 2004; Brülhart and Sbergami, 2009; Combes, Duranton, Gobillon, and Roux, 2010). And there might be two channels that agglomeration could affect productivity. First, wage level effects imply that productivity is affected immediately after entering clusters, and alters the direction of influence after leaving clusters. Second, productivity increases because of greater on-the-job skill accumulation in agglomeration, which is usually referred to the wage growth effects (Winters, 2013).

Therefore, if the substitution effect dominates, higher income from agglomeration will increase work intensity. If the income effect dominates, higher income in agglomeration will decrease hours worked. Conversely, if workers earn higher (lower) income in clusters, and they work more (less) in clusters, it must be the substitution effect that dominates. Similarly, if workers earn higher (lower) income in clusters, but they work less (more) in clusters, it is more likely that the income effect dominates. This model predicts that salaried workers may behave in similar ways as the self-employed in clusters if all workers have similar preferences, which cannot be explained by the urban rat race effect.

3 Empirical Framework

This paper is trying to establish the causal relationship between agglomeration and hours worked of the self-employed. As discussed in Section 2, there are three main channels which the causal relationship may run through, i.e., competition within industries, agglomeration wage effect, and endogeneity. Firstly, as mentioned above, our localization measure could pick up the competition within industries. Second, two instrumental variables are employed to take care of the sorting and simultaneity issue, and the causal relationship is also established by the instrumental variable estimation strategy. Lastly, although it is well documented in the literature, it is still necessary to revisit the agglomeration wage effect in our sample, since we will control for the agglomeration wage effect in the hours worked test.

3.1 Competition and Specialization

The relationship between agglomeration and hours worked will be estimated by Ordinary Least Squares (OLS) first. The baseline model is:

$$\log(y_{icd}) = \alpha \log(Urbanization_c) + \beta \log(Localization_{cd}) + \boldsymbol{X}_{icd}\boldsymbol{\gamma} + \mu_{icd} \quad (1)$$

where i indexes individual observations, c denotes work PUMA, and d denotes industries.[1] The dependent variable is log of hours worked. $Urbanization_c$ is the population density of a work PUMA. $Localization_{cd}$ is measured by the industry-specific employment density of a work PUMA.[2] $\boldsymbol{X}_{icd}$ is a standard set of demographic characteristics.

In the early version of this study, urbanization and localization are used alternatively in the regression. The results show that localization measure actually captures urbanization economy if it is solely included in the regression.[3] Therefore, collinearity may arise if urbanization and localization are controlled for simultaneously.[4] In order to mitigate the collinearity issue and obtain the net (real) effect of localization, the localization measure in Model (1) is replaced by the quotient of localization and urbanization. According to the intuition of this term, it is better to call the quotient relative localization.[5] The first model becomes to:

$$\log(y_{icd}) = \alpha \log(Urbanization_c) + \beta \log\left(\frac{Localization_{cd}}{Urbanization_c}\right) + \boldsymbol{X}_{icd}\boldsymbol{\gamma} + \epsilon_{icd} \qquad (2)$$

[1] All the empirical steps including all the regressions use personal sampling weights (perwt in IPUMS) to ensure the results are nationally representative.
[2] To construction the localization measure, the employment in each industry is calculated for each work PUMA. Then using this industry-specific employment divide by the geographic area of the work PUMA.
[3] The results are provided upon request.
[4] This is evident by the sophisticated correlation between urbanization and localization. The partial R squared is 0.4894. Thus, the partial correlation is about 0.7, implying highly collinearity may exist.
[5] Notice that although the definitions of localization and relative localization are different, relative localization better identifies the localization effect than the original localization measure. Thus, I will use "localization" as referring to "relative localization" in the remainder of this paper when discussing the effect of $\log\left(\frac{Localization_{cd}}{Urbanization_c}\right)$.

The OLS estimates potentially suffer from the omitted variable bias. Positive productivity shocks at the work PUMA level may attract more entrepreneurs, while concurrently affecting the work intensity. Amenity shocks could also affect agglomeration and work intensity. Under the spatial equilibrium assumption, productivity and amenity attractiveness are capitalized into wages and rents (Roback, 1982). Considerable studies document the potential endogeneity for including wages and rents directly (Rosenthal and Strange, 2008; Winters, 2013; Kahn and Lang, 1991). The reduced-form approach (Rosenthal and Strange, 2008; Kahn and Lang, 1991) provides little evidence of appropriately correcting the bias. Winters (2013) includes regression-adjusted wages and rents in the specifications to control for the productivity and amenity shocks. However, controlling for regression-adjusted wages could also capture the agglomeration wage effect. Therefore, amenities are directly included in the model, including violent crime, property crime, precipitation, January temperature, July temperature, elevation, minimum distance to the nearest river or lake, heating degree days, cooling degree days, dew points, direct solar irradiance, and four dummies for coastal work PUMAs of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes. All amenities are measured at the work PUMA level. Besides, since workers are more likely living outside of the work PUMAs, rents are not necessarily included in the specifications. The work MSA fixed effects mentioned below could take care of the characteristics of an area, including average rents.

In our preferred specification, industry and MSA fixed effects are also included as well. Considering it is possible that some workers are working in a MSA, but living in another MSA.[6] We also include residential MSA fixed effects as a robustness check. The identification here comes from those commuters who work in a MSA but live in another MSA or non-metropolitan area. The preferred specification is:

$$\log(y_{icdm}) = \alpha \log(Urbanization_c) + \beta \log\left(\frac{Localization_{cd}}{Urbanization_c}\right) + \boldsymbol{X}_{icdm}\boldsymbol{\gamma} + \boldsymbol{A}_c\boldsymbol{\theta} + \tau_d + \pi_m + \varepsilon_{icdm} \tag{3}$$

where *m* denotes work MSAs. $\boldsymbol{A}_c$ is a set of amenities. τ_d is industry fixed effect. π_m is work MSA fixed effect. ε_{icdm} is a white noise.

3.2 Sorting and Simultaneity

Admittedly, the endogeneity issue still could not be eliminated after controlling for a large set of observed variables. As discussed in Section 2, it is possible that people who like working longer hours may sort into denser or more competitive areas by some unobserved characteristics. Another interesting channel is a reverse relationship that more hours worked may cause more competition. To take care of sorting and simultaneity, instrumental variable estimation is employed. Minimum distance between work PUMA centroid and the USA shoreline[7] and estimated industry share in 1930 are used as the instrument variables in this paper.

To have a preview of our instruments, we have a preliminary discussion on the relevance and exogeneity conditions respectively. The relevance condition requires a valid instrumental variable is strongly correlated to the instrumented variable. Looking

[6] Most workers live in and work in the same MSA (84.30% self-employed in the sample live and work in the same MSA).
[7] It includes the shorelines of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes.

back to the American history, the first group of immigrants arrived in the north eastern coast of American continent in 1620. Then they began the history of the US from the coast to the inner land. However, because of the hardship in early days and resistance of Native Americans, westward expansion was very slow, and population accumulated in the eastern coast. Till 19th century, the Gold Rush brought another migration wave from coast to coast, settling down in California and Oregon Territory. The discovery of gold pulled rapid population growth in the western coastal areas. Figure 1.1 shows the historical expansion of the US.

Figure 0.1: A map of the historical territorial expansion of the United States of America. Source: National Atlas of the United States.

Before the invention of airplanes, the only way for immigrants to arrive in the US was crossing the oceans, which naturally caused the higher population in coastal areas than the hinterland. Furthermore, the increasing international trade kept the prosperity of ports and harbors. More recently, people treat beach areas as a great amenity. All these factors make coastal areas much denser, which is still true according to Figure 1.2 showing the population density for each work PUMA in 2000.[8]

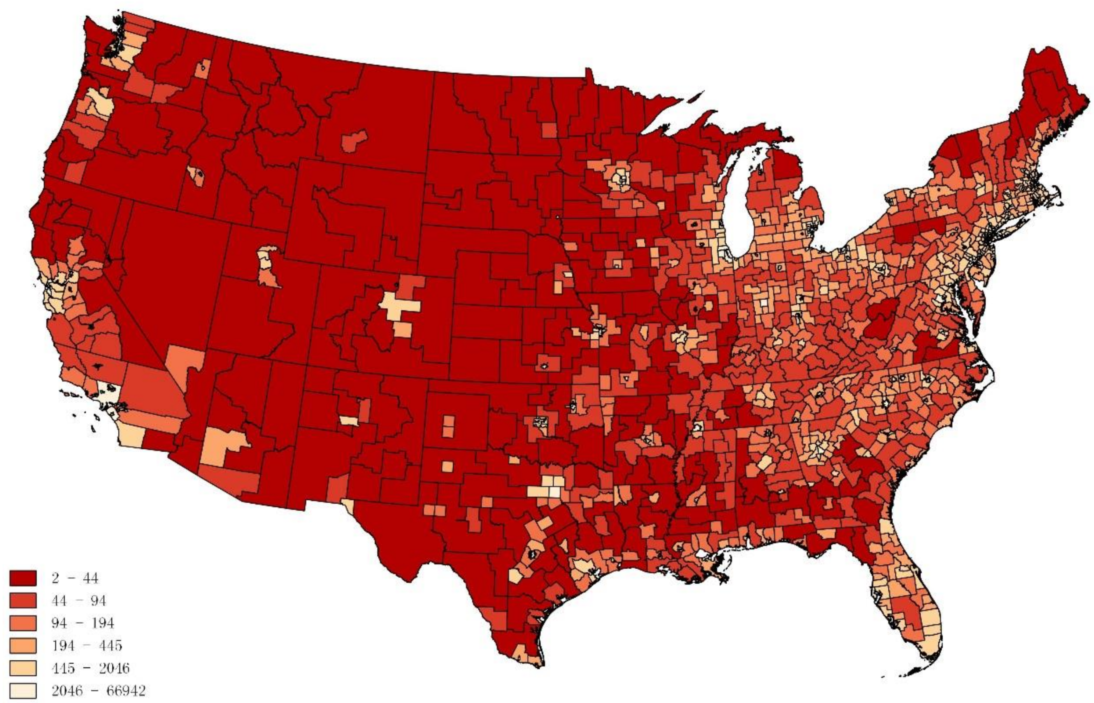

Figure 0.2: 2000 Population density at work PUMA level. (The magnitudes are classified by quantiles.) Source: Author.

By now, it is shown the reasonability of the correlation between population concentration and distance to coasts. The instrument in this paper is shown in Figure 1.3. We calculate the minimum distance from work PUMA centroid to shoreline (in red), which are the lengths of those orange lines. Different colors of work PUMAs indicate the distance differentials. Darker color indicates further distance to shoreline, lighter color

---

[8] According to our definition, Figure 1.2 indicates the magnitudes of urbanization measure. Unfortunately, we could not graphically present localization measure in the figure, since localization is an industry-specific measure at sub- work PUMA level.

indicates closer distance to shoreline. Comparing Figure 1.3 with Figure 1.2, we could find a similar pattern.

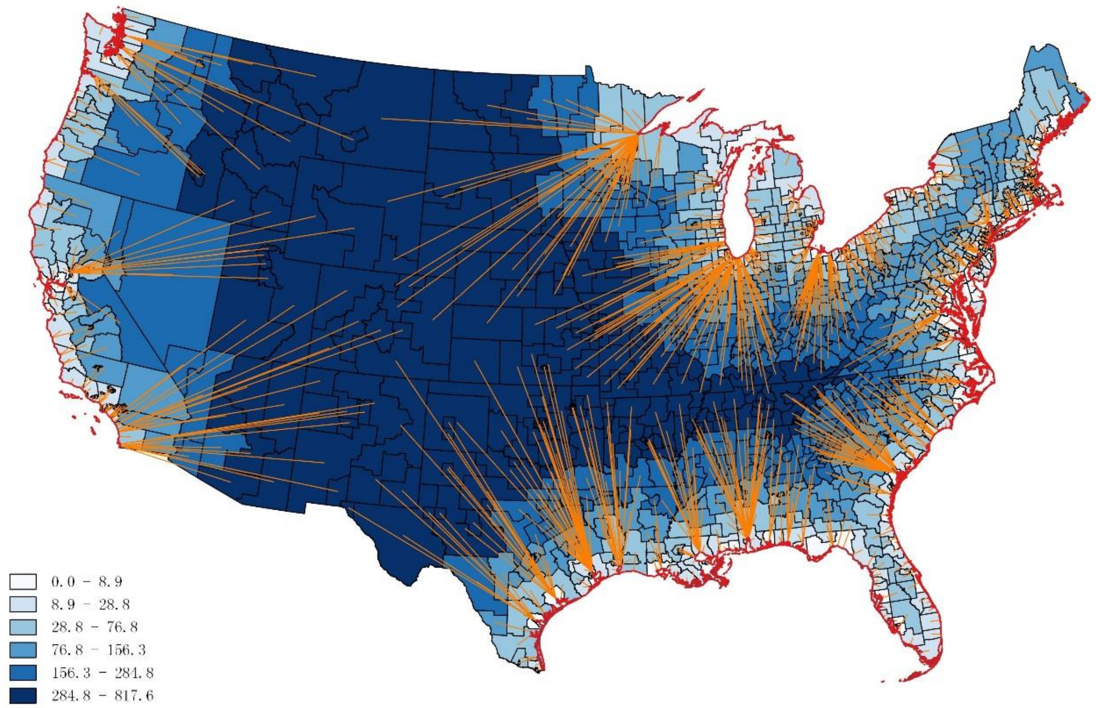

Figure 0.3: Minimum distance from work PUMA centroid to coastline. (The magnitudes in miles are classified by quantiles.) Source: Author.

The second instrument, estimated industry share in 1930, is arguably reliable since the historical (long lagged) variables are often used in the literature as instrumental variables. For example, a popular instrument for population density is historical population density (Ciccone and Hall, 1996; Combes et al., 2010). It is believed that the historical variables are relatively exogenous to current economic outcomes. To impute industry shares in 1930 at the work PUMA level, the employment by industry in 1930 is calculated at county level.[9] Then the county level data is converted to work PUMA level

[9] The data is from IPUMS 1930 5% sample. IND1950 is available in this sample, which identify industries by 1950 basis. Because IND1950 is also available in IPUMS 2000 sample, the consistency is guaranteed.

by using the allocation factor from MABLE/Geocorr2K: Geographic Correspondence Engine with Census 2000 Geography available at Missouri Census Data Center, which is used as the estimated allocation factor in 1930. Lastly, industry shares are calculated for each work PUMA. We expect the second instrument is strongly correlated with the relative localization measure.

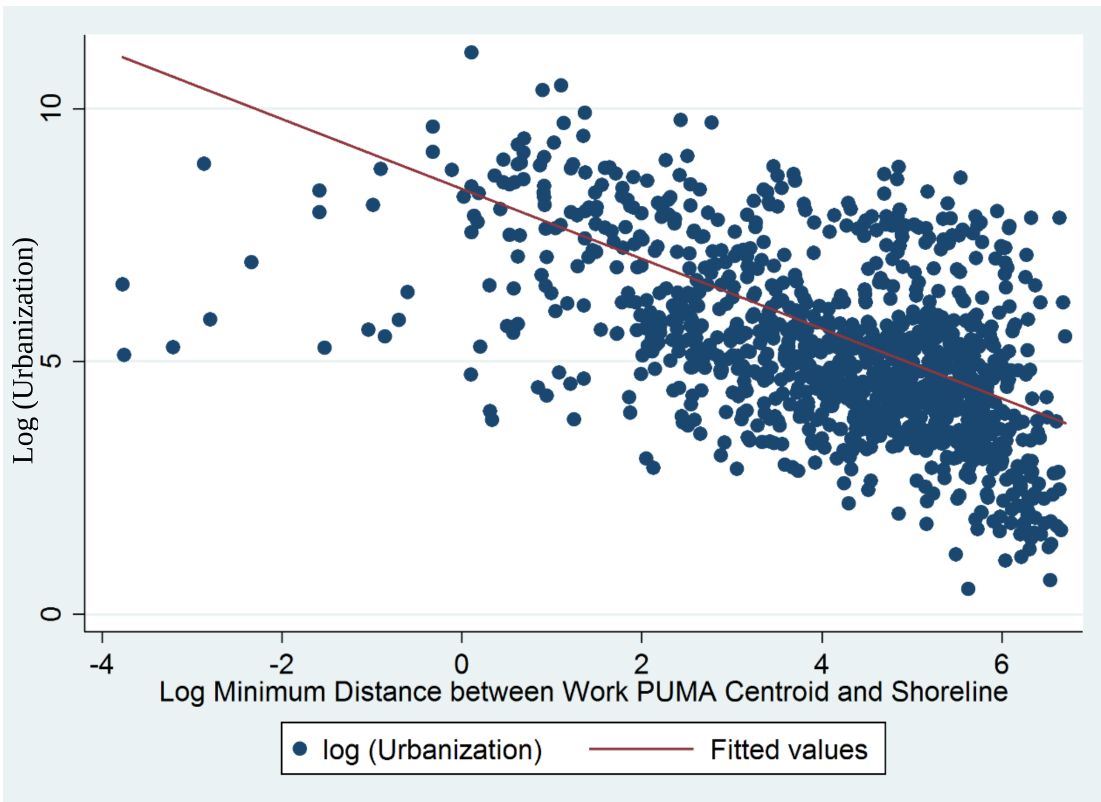

Figure 0.4: Raw correlation between log urbanization and log minimum distance to shoreline. Source: Author.

To have a more formal test, the correlation coefficients between log population density and log minimum distance to shoreline, and between log adjusted relative localization and log industry share in 1930 are calculated. Figures 1.4 and 1.5 show the raw correlation coefficients are -0.6051 and 0.6160 respectively, which are relatively strong associations. Besides, the formal first stage weak identification tests in Section 5 will show that our instruments are less likely suffering from the weak instrument issue.

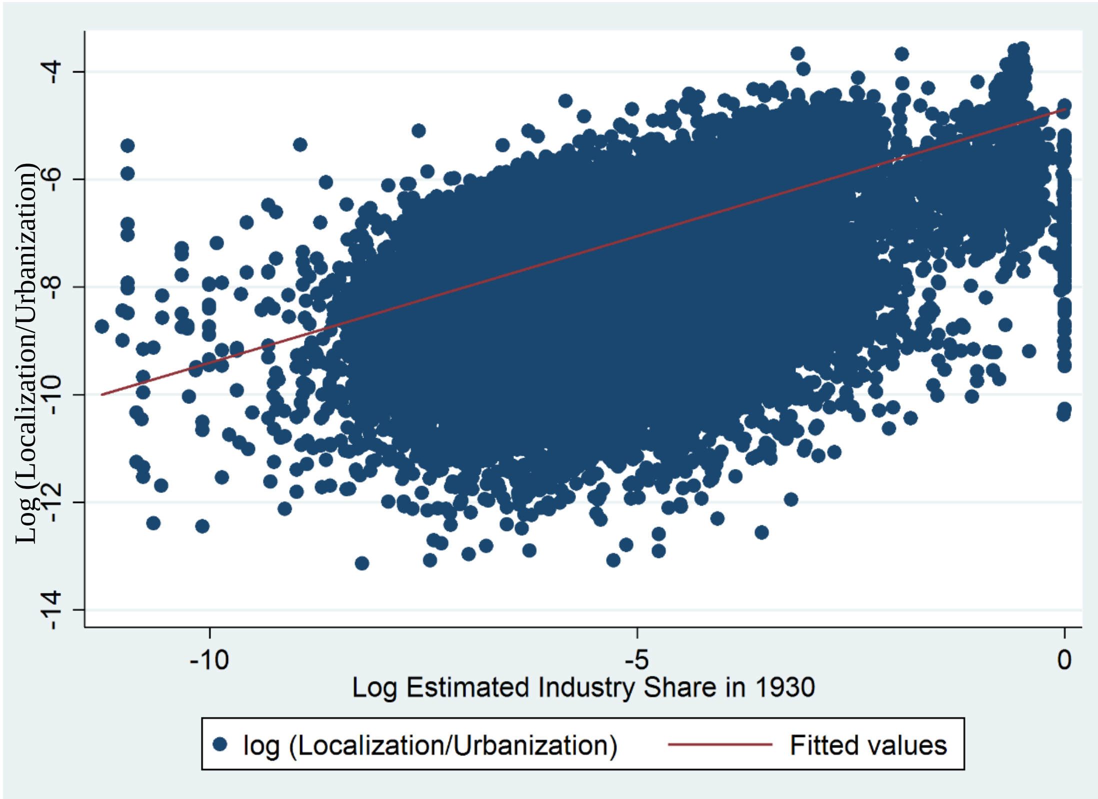

Figure 0.5: Raw correlation between log localization and log estimated industry share in 1930. Source: Author.

The exogeneity condition requires that a valid instrument is not causally related to the dependent variable in the second stage. In other words, taking a look at our first instrument, minimum distance to shoreline can only affect hours worked through agglomeration, and it seems to be the case. However, considering the agglomeration wage effect test in the next section, one could argue that minimum distance to shoreline affects productivity because it actually measures the minimum distance to ports and harbors, which could increase productivity. To show this concern is less important, it should be shown that ports and harbors are not correlated with productivity. One possible solution is directly controlling for the work PUMAs with ports and harbors. In our preferred specification, four dummies for coastal work PUMAs are controlled for, which includes the work PUMAs with ports and harbors. Another strategy is excluding all coastal work PUMAs from our sample. Section 5 will show the results as a robustness check.

To our knowledge, minimum distance to shoreline is rarely being used to instrument for the distribution of population. In the literature, the most related is soil quality used as instrument for population density by Combes et al. (2010) in a similar context. However, their identification suffers from weak instrument issue, even though they have several sub- measures of soil. Furthermore, they believe that Limited Information Maximum Likelihood (LIML) is weak instrument robust. However, LIML is not robust when we experiment a weak instrument in this paper.[10]

3.3 Revisit Agglomeration Wage Effect

3.3.1 Construct the Wage Measure

In order to examine the agglomeration wage effect, a reasonable measure of wage should be constructed. Regression-adjusted hourly incomes are an appropriate option, which are computed as the work PUMA fixed effects from the model:

$$\ln(Hourly\ Income)_{icd} = \boldsymbol{X}_{icd}\boldsymbol{\beta} + \tau_d + \omega_c + \varepsilon_{icd} \tag{4}$$

where τ_d is industry fixed effects. ω_c denotes the regression-adjusted average log hourly income in a work PUMA. In order to obtain a more exogenous wage measure, this regression is run for the self-employed and the employed separately. Then we will get

[10] Average elevation of work PUMA is experimented as the instrument for agglomeration measures. The first stage F statistics indicates it is a weak instrument. Then LIML is used to estimate the model. However, the bias is quite substantial.

regression-adjusted hourly income $\omega_c^{Self-Employed}$ for the self-employed, and $\omega_c^{Employed}$ for the employed. In the agglomeration wage effect test for the self-employed, $\omega_c^{Employed}$ is used as the welfare measure.[11] It is because controlling for $\omega_c^{Employed}$ in the regressions could capture the spillover aspect of agglomeration, and could eliminate any mechanism endogeneity in the hours worked regressions for the self-employed.

3.3.2 Agglomeration Wage Effect

Firstly, Ordinary Least Squares (OLS) is employed to estimate the agglomeration wage effect. The regression model is similar as the preferred specification of the hours worked test:

$$\omega_c^{Employed} = \alpha \log(Urbanization_c) + \beta \log\left(\frac{Localization_{cd}}{Urbanization_c}\right) + \boldsymbol{X}_{icdm}\boldsymbol{\beta} + \boldsymbol{A}_c\boldsymbol{\gamma} + \tau_d + \pi_m + \epsilon_{icdm} \qquad (5)$$

Admittedly, it is impossible to include all related variables. Even though one can include all the observable variables, there might be still some unobserved variables. Besides, the agglomeration measures may suffer from the measurement error bias. In order to take care of these issues and establish the causal relationship running from agglomeration to wages, the same instruments are used as in the hours worked test, i.e. minimum distance from work PUMA centroid to shoreline and estimated industry share in 1930. Minimum distance from work PUMA centroid to shoreline is possibly less valid

[11] As a robustness check, $\omega_c^{Self-Employed}$ is also used as the alternative welfare measure.

in this context as we discussed above. However, the robustness checks in section 5 will guarantee that it is valid in our sample.

The last important question is if the causal relationship between agglomeration and work intensity is driven by the agglomeration wage effect. To test this mechanism, the regression-adjusted incomes, $\omega_c^{Employed}$, are included in the hours worked test regression.

4 Data and Variables

4.1 Sample and Demographic Controls

This paper uses 5% IPUMS 2000 sample covering the contiguous 48 states. Only male full-time workers aged from 30 to 59[12] are included who work for 35 hours or more per week. In order to understand the age dimension of the relationship, the sample is subdivided into three groups: young group is aged from 30 to 39, middle-aged is between ages 40 and 49, and old workers are between 50 and 59.

For each subsample, they are further divided into two educational groups: high school degree or less, and college degree or more. College dropouts are excluded from the sample in order to ensure the division is sharp.[13]

[12] People aged 30 - 59 cover about 80% of the whole sample.

[13] College dropout are a very special group of people from the other two groups. They cannot be integrated into any other groups due to different behavior patterns. The empirical results show that most estimates for college dropouts are trivial, which can be provided upon request.

The class of worker (classwkr) variable is used to identify the self-employed. And the detailed variable of the class of worker (classwkrd) is used to identify entrepreneurs and other business owners as incorporated and unincorporated. The people either working in non-MSAs and/or living in non-MSAs are included in the sample by recoding the identifier with the work/residential state variables (pwstate2/statefip).[14]

All estimated models in this paper control for a standard set of demographic attributes, including educational attainment, a dummy of the presence of children, dummies of marital status, polynomial of age, dummies of race, years of residency in the US, and travel time to work.

4.2 Agglomeration Measures

In order to calculate the population density for a work PUMA, population and land area data of work PUMA is needed. Unfortunately, the data are only available for residential PUMAs.[15] Since work PUMA is coded differently from residential PUMA, population and land area data of work PUMA is calculated from corresponding residential PUMA by matching their codes.[16]

Localization measure is constructed within the sample. Employment in each industry is calculated for each work PUMA adjusted by personal weight. Then using this industry-specific employment divide by the geographic area of the work PUMA to obtain localization measure. The variable of IND1950 in the IPUMS is used to identify

---

[14] The sample excluding people work in non-MSA is also tested, the results are omitted due to similarity.
[15] Population and land area data for 2000 work PUMA are extracted from Missouri Census Data Center at http://mcdc.missouri.edu/websas/geocorr2k.html
[16] The relationship between residential PUMA and work PUMA is provided by the IPUMS table: https://usa.ipums.org/usa/volii/00pwpuma.shtml#5percent

industries, which is a three-digit identifier. 124 three-digit industry categories are used to construct the localization measure.[17] Then we recode IND1950 to a two-digit identifier based on broader groups. 11 two-digit industry categories are included as industry fixed effects.

4.3 Dependent Variables

Since people may have multiple sources of employment, INCEARN is used as the annual income measure, which is the sum of wage income, business income, and farm income in the previous year. Then hourly incomes are obtained by dividing the annual incomes by the hours worked in the previous year for each observation.

This paper uses usual hours worked per week in the previous year (uhrswork in IPUMS) as the measure of work intensity.[18] Kahn and Lang (1991) find that using actual hours worked rather than desired hours for the self-employed will not cause bias. This bias is caused by the deviation between actual and desired hours, since the self-employed are less restricted to choose their hours worked, and they are well compensated for working longer hours.

4.4 Amenities

Amenities are extracted from different sources and constructed at the work PUMA level, including violent crime, property crime, precipitation, January temperature, July

---

[17] All salaried workers are excluded.

[18] I also experiment with annual hours worked. In order to get the annual data, I use hours worked per week multiply by weeks worked (wkswork1 in IPUMS). The results are similar. But there are some concerns of the annual data. As the description from IPUMS, "For employers, WKSWORK1 covers all weeks that the business or farm was in operation, even if the employer was absent." And thus, this data is not actual hours, and there might be large measurement error.

temperature, elevation, minimum distance to the nearest river or lake, heating degree days, cooling degree days, dew points, direct solar irradiance, and four dummies for coastal work PUMAs of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes.

Crime data comes from the Uniform Crime Reporting Program Data (UCR). This paper uses its county-level detailed arrest and offense data in 2000, which covers all counties in the US except for Wisconsin, Illinois, DC, and Florida. Violent crime and property crime data for Wisconsin, DC, and Florida come from USA Counties Website, but the Illinois data is still missing. Crime data for Illinois counties is extracted from Illinois County Website. This county-level data is converted to PUMA level by using the allocation factor from MABLE/Geocorr2K: Geographic Correspondence Engine with Census 2000 Geography available at Missouri Census Data Center. Then the geocodes for PUMA are recoded for work PUMA.

Precipitation and dew points data are obtained from the old version of PRISM.[19] They are 30-arc-second (800 meters) gridded raster data. For precipitation, 30-year (1971 – 2000) annual average data is used. For dew point, 10-year (1991 – 2000) annual average data is constructed. Boundary file for work PUMAs is available at IPUMS website.[20] Then mean precipitation and dew points data for each work PUMA are calculated by GIS software.

---

[19] PRISM Climate Group, Oregon State University, http://oldprism.nacse.org, created 19 Sep 2015.
[20] https://usa.ipums.org/usa/volii/00pwpuma.shtml

January temperature and July temperature are extracted from current version of PRISM.[21] Monthly raster data at 4 kilometer grid cell resolution from 1981 – 2000 is used to construct the 20 year average data. Then mean January temperature and July temperature for each work PUMA are calculated using the work PUMA shapefile.

Elevation data are extracted from the Hole-filled Seamless SRTM data V4.1 distributed by the International Centre for Tropical Agriculture (CIAT). The data source is Shuttle Radar Topography Mission (SRTM) of National Aeronautics and Space Administration (NASA) available at U.S. Geological Survey (USGS). The original SRTM data are available at 1 arc-second and 3 arc-second grid cell resolutions but with small voids. The data distributed by CIAT filled the voids using interpolation methods with 3 arc-second grid cell resolution (approximately 90 meters). The Global 30 Arc-Second Elevation (GTOPO30) data are also experimented, the results are similar. Considering its lower resolution, results from GTOPO30 will not be presented in this paper. Average elevation for each work PUMA is calculated using the shapefile.

River centerlines and lakes shapefiles with 1:10 million scale are available at Natural Earth website.[22] The global datasets are merged with North America supplement datasets. River centerlines and lakes in the US are clipped from the global merged data by work PUMA shapefile. Great Lakes are excluded from the dataset, since they are used to construct the coastline. Minimum distance from work PUMA centroid to nearest river or lake is calculated by GIS software.

[21] PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 20 Sep 2015.
[22] http://www.naturalearthdata.com/downloads/10m-physical-vectors/

Solar irradiance, heating degree days, and cooling degree days are retrieved from National Renewable Energy Laboratory (NREL) under the U.S. Department of Energy. Direct Normal Irradiance (DNI) at 10 kilometer resolution for lower 48 States is used to constructed the average data for each work PUMA. Heating degree days and cooling degree days are derived by Solar and Wind Energy Resource Assessment (SWERA) from NASA Surface meteorology and Solar Energy (SSE) dataset. One-degree cell resolution GIS data are available at the NREL website. Then a similar approach is applied to construct the work PUMA level data.

Four dummies for coastal work PUMAs of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes are constructed using the shapefile of 2000 work PUMA. The coastline is derived from the shapefile of work PUMA as well to ensure consistency. If a work PUMA shares its boundary with any one of the four coastlines, it will be assigned a value of one for the corresponding coastal work PUMA dummy. Zeros are assigned to those work PUMAs not attached to any of the coastlines.

Table 1.2 shows the summary statistics for the self-employed in our sample. There are 210,910 weighted observations in the sample.

Table 1.2: Summary Statistics

Variable	No. Obs	Mean	Std. Dev.	Min	Max
Hourly Income	210,910	26.328	47.574	-285.714	5485.714
Log (Hourly Income)	206,547	2.683	1.168	-7.160	8.610
Hours Worked	210,910	50.076	11.936	35.000	99.000
Log (Hours Worked)	210,910	3.888	0.219	3.555	4.595
Localization	210,910	9.322	55.907	0.000	716.980
Log (Localization)	210,910	-0.727	2.223	-9.757	6.575
Urbanization	210,910	3234.938	10826.370	1.637	66942.260
Log (Urbanization)	210,910	5.933	2.012	0.493	11.112
Minimum Distance to Coastline	210,910	147.952	182.249	0.023	817.575
Log (Minimum Distance to Coastline)	210,910	3.912	1.824	-3.772	6.706

Imputed Industry Share in 1930	210,910	0.064	0.147	0.000	1.000
Log (Industry Share in 1930)	210,910	-4.206	1.659	-11.259	0.000
High School and Less	210,910	0.535	0.499	0.000	1.000
College and More	210,910	0.465	0.499	0.000	1.000
Age	210,910	44.718	7.874	30.000	59.000
Log (Commute Time)	176,068	2.770	0.982	0.000	5.159
Children Present	210,910	0.569	0.495	0.000	1.000
Marital Status					
Married	210,910	0.751	0.432	0.000	1.000
Married, Spouse Absent	210,910	0.012	0.109	0.000	1.000
Separated	210,910	0.017	0.129	0.000	1.000
Divorced	210,910	0.113	0.317	0.000	1.000
Widowed	210,910	0.010	0.101	0.000	1.000
Never Married	210,910	0.096	0.295	0.000	1.000
Race					
White	210,910	0.867	0.340	0.000	1.000
African American	210,910	0.039	0.195	0.000	1.000
American Indian or Alaska Native	210,910	0.005	0.069	0.000	1.000
Chinese	210,910	0.012	0.111	0.000	1.000
Japanese	210,910	0.002	0.050	0.000	1.000
Other Asian or Pacific Islander	210,910	0.030	0.171	0.000	1.000
Other Race	210,910	0.028	0.165	0.000	1.000
Two Major Races	210,910	0.016	0.124	0.000	1.000
Three or More Major races	210,910	0.001	0.026	0.000	1.000
Hispanic Origin					
Not Hispanic	210,910	0.931	0.253	0.000	1.000
Mexican	210,910	0.037	0.190	0.000	1.000
Puerto Rican	210,910	0.003	0.059	0.000	1.000
Cuban	210,910	0.005	0.071	0.000	1.000
Other	210,910	0.023	0.149	0.000	1.000
Amenity					
Log (Violent Crime)	210,587	6.670	1.713	1.946	10.380
Log (Property Crime)	210,673	7.850	1.526	2.639	12.000
Log (Precipitation)	210,910	8.943	0.465	6.677	9.998
Log (Dew Points)	210,910	7.147	0.716	-10.735	7.855
Log (January Temperature)	210,910	2.742	0.640	-13.356	3.540
Log (July Temperature)	210,910	3.175	0.143	2.608	3.499
Log (Heating Degree Days)	210,910	7.536	0.864	3.664	8.616
Log (Cooling Degree Days)	210,910	7.659	0.431	6.157	8.556
Log (Elevation)	210,910	8.407	0.805	-7.953	8.904
Log (Solar Irradiance)	210,910	1.515	0.203	1.129	2.069
Log (Minimum Distance to River and Lake)	210,910	2.159	1.203	-2.052	4.432
Atlantic Work PUMA	210,910	0.157	0.363	0.000	1.000
Great Lake Work PUMA	210,910	0.052	0.223	0.000	1.000
Gulf Work PUMA	210,910	0.051	0.220	0.000	1.000
Pacific Work PUMA	210,910	0.129	0.335	0.000	1.000

Notes: All summary statistics are adjusted by personal weight to ensure the national representative. Education, years of residency in the U.S., industry, work PUMA, and work MSA are not included for space conservation.

5 Empirical Results

5.1 Hours Worked and Agglomeration

In order to have a clear idea on the mechanism that agglomeration could affect work intensity, this paper uses OLS to test the descriptive correlation first. Then 2SLS will be employed to examine the sorting and simultaneity. Lastly, the agglomeration wage effect will be included to test whether agglomeration affects work intensity through wage differentials.

Panel A of Table 1.3 reports the OLS estimates of the regression model (3).[23] Surprisingly, nearly all estimates for urbanization are negative. Especially for the lower educated group, the estimates are quite significant and substantial. Although we do not expect competition across industries, it is somewhat shocking to have a negative relationship between urbanization and hours worked. On the other hand, all the estimates of relative localization are positive and significant as expected. The positive relationship means the self-employed work more hours in more competitive areas, which is also documented by Rosenthal and Strange (2008).

Panel B of Table 1.3 shows the 2SLS results with minimum distance from work PUMA centroid to coastline and estimated industry share in 1930 as the instruments. All estimates increase in their magnitudes because of the potential measurement errors in the two agglomeration measures. The negative association between urbanization and hours

[23] The estimates of detailed controls are provided upon request.

Table 1.3: Hours Worked and Agglomeration

Dependent Variable: Log (Hours Worked)						
	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
A. Ordinary Least Squares						
Log (Urbanization)	-0.0055**	-0.0080***	-0.0049*	-0.0010	0.0001	-0.0087***
	(0.0028)	(0.0023)	(0.0027)	(0.0036)	(0.0028)	(0.0028)
Log(Localization/Urbanization)	0.0224***	0.0118***	0.0123***	0.0077***	0.0075***	0.0102***
	(0.0028)	(0.0023)	(0.0023)	(0.0023)	(0.0018)	(0.0019)
B. Two Stage Least Squares						
Log (Urbanization)	-0.0089	-0.0087	-0.0114	0.0024	-0.0179*	-0.0191**
	(0.0128)	(0.0089)	(0.0128)	(0.0107)	(0.0101)	(0.0080)
Log(Localization/Urbanization)	0.0391***	0.0395***	0.0593***	0.0199***	0.0266***	0.0281***
	(0.0072)	(0.0067)	(0.0083)	(0.0075)	(0.0055)	(0.0059)
First Stage						
Urbanization						
Log (Distance to Shoreline)	-0.2833***	-0.2970***	-0.2668***	-0.5017***	-0.4744***	-0.4563***
	(0.0579)	(0.0558)	(0.0545)	(0.0664)	(0.0613)	(0.0610)
Log (Industry share in 1930)	0.0410***	0.0489***	0.0411***	0.0592***	0.0560***	0.0472***
	(0.0108)	(0.0098)	(0.0094)	(0.0109)	(0.0100)	(0.0097)
Localization/Urbanization						
Log (Distance to Shoreline)	-0.0270	-0.0136	-0.0400	-0.2213**	-0.1928**	-0.1615***
	(0.0315)	(0.0301)	(0.0305)	(0.0888)	(0.0751)	(0.0620)
Log (Industry share in 1930)	0.2748***	0.2821***	0.2701***	0.3477***	0.3446***	0.3322***
	(0.0166)	(0.0164)	(0.0152)	(0.0210)	(0.0191)	(0.0194)
Underidentification	18.823	25.106	23.207	39.280	42.027	41.468
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Weak Identification	13.032	16.312	14.517	42.284	45.662	45.221
	{7.03}	{7.03}	{7.03}	{7.03}	{7.03}	{7.03}
Endogeneity	6.300	24.692	34.647	6.021	12.347	9.956
	[0.0429]	[0.0000]	[0.0000]	[0.0493]	[0.0021]	[0.0069]
Agglomeration Wage Effect	No	No	No	No	No	No
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 1.2. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. P-values in square brackets are provided for underidentification tests and endogeneity tests. Stock-Yogo weak identification test critical values in braces are provided for weak identification test. * p < 0.1, ** p < 0.05, *** p < 0.01.

worked becomes less significant by using instrument variable estimation, which is closer to the expectation that the rivalry is trivial across industry. Meanwhile, the estimates for relative localization are still significant. Localization causes 3.91% - 5.93% increase in hours worked for the less educated self-employed, and 1.99% - 2.81% increase for the higher educated group. Therefore, the instrumental variable estimation results imply that the positive relationship between localization and hours worked by the self-employed is less likely resulted from sorting and simultaneity, but directly comes from competition within industries.

The significant and considerable first stage estimates indicate that minimum distance to shoreline is negatively correlated to agglomeration, and historical industry share is positively associated with agglomeration. The large first stage Kleibergen-Paap Wald F statistics provide solid evidence that our instruments are less likely to suffer from the weak instrument issue.

Now, we further divide our sample into the incorporated self-employed and the not incorporated self-employed. The debate on whether entrepreneurship returns still exists in the literature, and the main debate is on the choice of proxy of entrepreneurship. Borjas and Bronars (1989), Evans and Leighton (1989), and Hamilton (2000) use the self-employed as a proxy for entrepreneurs, concluding that entrepreneurship does not pay. However, using aggregate self-employed as a proxy for entrepreneurs makes "little distinction between Michael Bloomberg and a hot dog vendor" (Glaeser, 2007). Entrepreneurs are naturally different from other self-employed individuals. Although Faggio and Silva (2014) argues that this difference is not important in urban areas, it is important to distinguish them for studies including rural areas. Levine and Rubinstein

(2012) separate the self-employed into incorporated and unincorporated to distinguish entrepreneurs and other business owners, and find that entrepreneurs earn much more than the other types of workers. Thus, this paper will follow the definition of entrepreneurs and other self-employed by Levine and Rubinstein (2012) to test if the relationship between agglomeration and hours worked is different for the sub-samples. However, the 2SLS results for the not incorporated self-employed and the incorporated self-employed are very similar to the self-employed as a whole.[24]

5.2 Agglomeration Wage Effect

Before we control for the agglomeration wage effect in the hours worked test to examine whether it is a possible mechanism, it is necessary to revisit the agglomeration wage effect first, which is well documented by the literature that population geographic concentration increases wages and productivity though (Ciccone, 2002; Rosenthal and Strange, 2004; Brülhart and Sbergami, 2009; Combes, Duranton, Gobillon, and Roux, 2010).

Panel A of Table 1.4 reports the agglomeration wage effect by OLS of Model (5). All estimates for urbanization are statistically significant and large in magnitudes, implying that urbanization is correlated with higher wages. Although the estimates for relative localization are still positive, the magnitudes and significance are much lower than those for urbanization. Only the estimates for higher educated group are significant at the 0.05 level.

[24] Results are provided upon request.

Table 1.4: Agglomeration Wage Effects

	Dependent Variable: Log (Adjusted Income for Employee)					
	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
A. Ordinary Least Squares						
Log (Urbanization)	0.0306***	0.0324***	0.0311***	0.0307***	0.0297***	0.0291***
	(0.0036)	(0.0036)	(0.0037)	(0.0046)	(0.0043)	(0.0041)
Log (Localization/Urbani:	0.0012	0.0009	0.0003	0.0042*	0.0039*	0.0038**
	(0.0017)	(0.0015)	(0.0014)	(0.0022)	(0.0020)	(0.0016)
B. Two Stage Least Squares						
Log (Urbanization)	0.0634***	0.0638***	0.0738***	0.0556***	0.0489***	0.0448***
	(0.0177)	(0.0158)	(0.0197)	(0.0177)	(0.0159)	(0.0141)
Log (Localization/Urbani:	-0.0004	-0.0014	-0.0051	-0.0013	-0.0006	0.0022
	(0.0034)	(0.0032)	(0.0037)	(0.0030)	(0.0028)	(0.0022)
First Stage						
Urbanization						
Log (Distance to Shoreline)	-0.2833***	-0.2970***	-0.2668***	-0.5017***	-0.4744***	-0.4563***
	(0.0579)	(0.0558)	(0.0545)	(0.0664)	(0.0613)	(0.0610)
Log (Industry share in 1930)	0.0410***	0.0489***	0.0411***	0.0592***	0.0560***	0.0472***
	(0.0108)	(0.0098)	(0.0094)	(0.0109)	(0.0100)	(0.0097)
Localization/Urbanization						
Log (Distance to Shoreline)	-0.0270	-0.0136	-0.0400	-0.2213**	-0.1928**	-0.1615***
	(0.0315)	(0.0301)	(0.0305)	(0.0888)	(0.0751)	(0.0620)
Log (Industry share in 1930)	0.2748***	0.2821***	0.2701***	0.3477***	0.3446***	0.3322***
	(0.0166)	(0.0164)	(0.0152)	(0.0210)	(0.0191)	(0.0194)
Underidentification	18.823	25.106	23.207	39.280	42.027	41.468
	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Weak Identification	13.032	16.312	14.517	42.284	45.662	45.221
	{7.03}	{7.03}	{7.03}	{7.03}	{7.03}	{7.03}
Endogeneity	11.536	13.114	13.831	5.494	3.166	4.718
	[0.0031]	[0.0014]	[0.0010]	[0.0641]	[0.2054]	[0.0945]
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 1.2. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. P-values in square brackets are provided for underidentification tests and endogeneity tests. Stock-Yogo weak identification test critical values in braces are provided for weak identification test. * p < 0.1, ** p < 0.05, *** p < 0.01.

Considering the endogeneity issue, Panel B of Table 1.4 presents the 2SLS results for the agglomeration wage effect. The first row shows the 2SLS estimates for log urbanization measure, which are not qualitatively different from the OLS estimates. However, comparing the magnitudes, the effect of agglomeration on wages is understated by the OLS. Especially for the lower educated self-employed, the downward bias is substantial, implying the measurement errors are influential. Comparing the 2SLS estimates with the OLS estimates, agglomeration increases wages by 3.06% - 3.24% for different age groups of the lower educated self-employed, but 6.34% - 7.38% increases are estimated by 2SLS. The agglomeration wage effects are about 2.91% - 3.07% for the highly educated self-employed by OLS, but the estimates increase to 4.48% - 5.56% by 2SLS. Meanwhile, all the relative localization estimates decrease in magnitudes and become insignificant. Most estimates even flip sign. Therefore, the agglomeration wage effect comes from urbanization rather than localization, which confirms the literature.

Note that, since the first stage estimations in Table 1.4 are identical to the first stage of hours worked test in Table 1.3, the first stage results are same as those in Table 1.3. Thus, our instrument does not suffer from weak instrument, and works well in the current context.

5.3 Controlling for Agglomeration Wage Effect

In order to test whether the agglomeration wage effect plays a role in the effect of agglomeration on work intensity, $\omega_c^{Employed}$ is included in the regressions. Table 1.5 reports the instrument variable estimation results.

Table 1.5: Hours Worked and Agglomeration

Dependent Variable: Log (Hours Worked)						
	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
Two Stage Least Squares						
Log (Urbanization)	-0.0139	-0.0116	-0.0211	0.0079	-0.0212	-0.0166
	(0.0191)	(0.0132)	(0.0196)	(0.0151)	(0.0137)	(0.0109)
Log (Localization/Urbaniz	0.0391***	0.0395***	0.0600***	0.0198***	0.0266***	0.0283***
	(0.0072)	(0.0068)	(0.0085)	(0.0075)	(0.0055)	(0.0059)
First Stage						
Urbanization						
Log (Distance to Shoreline)	-0.1838***	-0.1922***	-0.1669***	-0.3546***	-0.3491***	-0.3396***
	(0.0589)	(0.0551)	(0.0556)	(0.0694)	(0.0665)	(0.0681)
Log (Industry share in 1930)	0.0272***	0.0338***	0.0327***	0.0440***	0.0422***	0.0312***
	(0.0092)	(0.0081)	(0.0080)	(0.0085)	(0.0084)	(0.0076)
Localization/Urbanization						
Log (Distance to Shoreline)	-0.0326	-0.0204	-0.0500**	-0.1711***	-0.1520***	-0.1241***
	(0.0248)	(0.0237)	(0.0240)	(0.0547)	(0.0466)	(0.0419)
Log (Industry share in 1930)	0.2756***	0.2831***	0.2709***	0.3425***	0.3401***	0.3271***
	(0.0161)	(0.0159)	(0.0147)	(0.0203)	(0.0186)	(0.0188)
Underidentification	7.347	9.766	7.464	16.307	17.046	15.998
	[0.0067]	[0.0018]	[0.0063]	[0.0001]	[0.0000]	[0.0001]
Weak Identification	4.679	5.908	4.352	13.600	15.248	14.551
	{7.03}	{7.03}	{7.03}	{7.03}	{7.03}	{7.03}
Endogeneity	5.971	24.561	34.034	6.146	12.237	9.886
	[0.0505]	[0.0000]	[0.0000]	[0.0463]	[0.0022]	[0.0071]
Agglomeration Wage Effect	Yes	Yes	Yes	Yes	Yes	Yes
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 1.2. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. P-values in square brackets are provided for underidentification tests and endogeneity tests. Stock-Yogo weak identification test critical values in braces are provided for weak identification test. ** $p < 0.05$, *** $p < 0.01$.

Compared with the corresponding results in Table 1.3, there is not much quantitative difference. The only change is that the urbanization estimates become strictly insignificant. Therefore, only localization increases work intensity of the self-employed mainly through competition within industries. The effect of sorting, simultaneity, and agglomeration wage effect are less important.

Nonetheless, considering the negative correlation between urbanization and hours worked, it mostly comes from endogeneity and agglomeration wage effect. After correcting these effects, urbanization does not appear to affect work intensity for the self-employed.

5.4 Robustness Checks

Before the final conclusions could be made, several robustness checks should be performed to reinforce the credibility of our study.

Firstly, as discussed in Section 3, our instrument is valid if we can show ports and harbors are not correlated with productivity. One possible solution is directly controlling for the work PUMAs with ports and harbors. Aggressively, in our preferred specification, four dummies for coastal work PUMAs are controlled for, which includes all the work PUMAs with ports and harbors. However, most estimates on the coastal dummies are insignificant.[25] Besides, our instruments are not weak instruments according to the weak identification statistics with the coastal dummies in these regressions.

Another strategy is excluding all coastal work PUMAs from our sample. Panel A of Table 1.6 reports the instrumental variable estimation results. Either for the agglomeration wage effect test or for the hours worked test, the estimates are not

[25] Results are provided upon request.

Table 1.6: Robustness Checks

	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
A. Exclusion of Coastal Work PUMAs						
Dependent Variable: Hours Worked						
Log (Urbanization)	-0.0350	0.0051	0.0586*	0.0119	0.0408*	0.0121
	(0.0321)	(0.0226)	(0.0337)	(0.0238)	(0.0222)	(0.0217)
Log(Localization/	0.0285***	0.0381***	0.0434***	0.0205*	0.0183**	0.0258***
Urbanization)	(0.0106)	(0.0098)	(0.0125)	(0.0113)	(0.0078)	(0.0085)
Dependent Variable: Adjusted Wages of Employees						
Log (Urbanization)	0.1342***	0.1209***	0.1355***	0.0760***	0.0769***	0.0908***
	(0.0433)	(0.0312)	(0.0365)	(0.0154)	(0.0170)	(0.0222)
Log(Localization/	-0.0095	-0.0113	-0.0124	-0.0066	-0.0061	-0.0023
Urbanization)	(0.0095)	(0.0080)	(0.0087)	(0.0046)	(0.0041)	(0.0043)
B. Incomes for Employers						
Dependent Variable: Hours Worked						
Log (Urbanization)	-0.0077	-0.0058	-0.0122	0.0066	-0.0189	-0.0182**
	(0.0142)	(0.0103)	(0.0147)	(0.0126)	(0.0118)	(0.0092)
Log(Localization/	0.0390***	0.0391***	0.0595***	0.0197***	0.0266***	0.0281***
Urbanization)	(0.0072)	(0.0068)	(0.0084)	(0.0075)	(0.0056)	(0.0059)
Dependent Variable: Adjusted Incomes of Employers						
Log (Urbanization)	0.0961**	0.1032***	0.1088***	0.0906**	0.0833**	0.0737**
	(0.0375)	(0.0345)	(0.0404)	(0.0374)	(0.0339)	(0.0308)
Log(Localization/	-0.0073	-0.0113	-0.0179**	-0.0051	-0.0066	-0.0043
Urbanization)	(0.0072)	(0.0070)	(0.0075)	(0.0063)	(0.0059)	(0.0048)

Notes: This table reports the 2SLS estimates. The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 1.2. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

qualitatively different from those in the full sample specifications. Therefore, our instruments are robust considering one of them might affect productivity through another potential channel.

Lastly, we consider the wage measure. There are at least two advantages to use adjusted wages of employees in our study. First, since agglomeration is a spillover effect, we can capture this by using the wages of people out of the sample. Second, it could eliminate any mechanism endogeneity since wages of the employed are not directly associated with work intensity of the self-employed. Panel B of Table 1.6 reports the results by using $\omega_c^{Self-Employed}$ instead of $\omega_c^{Employed}$. The upper part presents the 2SLS estimates for hours worked test with controlling for $\omega_c^{Self-Employed}$. The results are similar to our main specification. The lower part shows the estimates for agglomeration wage effect test by using adjusted income of the self-employed as the dependent variable. The estimates are larger than those in Table 1.4, which is reasonable if considering the spillover effect of agglomeration.

6 Conclusions

This paper uses minimum distance from work PUMA centroid to the US coastlines and estimated industry share in 1930 as instrumental variables, and tests the causal impact of agglomeration on work intensity of the self-employed. The instrument variable estimation results show that localization has positive causal impact on work intensity of the self-employed. The causality does not exist for urbanization and hours worked controlling for agglomeration wage effect.

The empirical results imply that the positive causality is mostly driven by competition within industries relative to sorting and simultaneity. The benefits brought by localization attract more firms from the same industry. Increasing competition within industries causes longer hours worked to survive. On the other hand, although urbanization enhances productivity, there is no basis for firms from different industries to compete. Therefore, urbanization less likely affects hours worked.

Based on the results that we get in this paper, we can understand the relationship between agglomeration and labor supply more comprehensively. Rosenthal and Strange (2008) find that localization affects hours worked for all full-time workers. This paper focuses on the self-employed sample, and gets a similar result. However, the mechanism of the relationship for employees and the self-employed could be different. Rosenthal and Strange (2008) look at the urban rat race effect as the mechanism, which is an individual competition effect. This paper uses relative localization to capture firm level competition. The firm level competition usually happens within industries, which could cause the self-employed to work more.

The literature documents several benefits for firms locating in clusters. This paper provides additional evidence that clusters could bring regional benefits to local communities, since our results suggest that localization could increase the size of economic activities. Local governments could create designated zoning areas for industries to promote the process of localization, so that the local areas could enjoy the benefits with firms as well.

This paper also examines whether the agglomeration wage effect is a mechanism. The empirical results indicate that the agglomeration wage effect comes only from urbanization. However, the agglomeration wage effect is not a crucial mechanism that the positive causality of agglomeration on hours worked of the self-employed runs through.

Our instruments are relatively new. The first stage statistics and the robustness checks guarantee our instruments are reliable, which could inspire us to construct similar instruments. For example, we could use the minimum distance from land-grant universities to PUMA centroids to instrument for human capital stock.

The future study will focus on theoretical modeling of the relationship between agglomeration and labor supply to substantiate the empirical results we have in this paper.

CHAPTER II

SELF-EMPLOYMENT DIFFERENTIALS AMONG FOREIGN-BORN STEM AND NON-STEM WORKERS

1 Introduction

The U.S. is a country built and developed by immigrants. It is valuable to understand how immigrants contribute to the economy of this country, then strategic immigration policies could be formed to maximize the benefits brought by immigrants. Many studies suggest that small businesses and foreign immigrants educated in science, technology, engineering, and mathematics (STEM) are the two drivers of economies (Acs and Armington 2004; Audretsch and Keilbach 2004; Audretsch et al. 2006; Bruce et al. 2009; Hunt and Gauthier-Loisselle 2010; Winters 2014; Peri et al. 2015). However, little about the connection of these two is known. Especially, pro-STEM immigration policies could be further endorsed if STEM graduates are more likely to be self-employed, but adjustments should be considered if they are not.

This paper uses the 1% Integrated Public Use Microdata Series (IPUMS, Ruggles et al., 2010) 2013 American Community Survey (ACS) sample to study the correlation between college majors and self-employment decisions for foreign college graduates in the U.S. The self-employment differential between aliens educated in STEM and non-STEM is most of our interest. The differentials across non-STEM majors are also tested for further reference. Moreover, income differentials for the employed and self-employed across majors are examined to offer some possibilities to understand the differences in self-employment decisions.

Surprisingly, the empirical results show that foreign STEM graduates are significantly less likely to be self-employed than their non-STEM counterparts. This association remains even when considering broad major groups, and detailed majors. The income differential results indicate that STEM graduates could earn significantly more than the other majors if they choose not to be self-employed, and their income will significantly drop if they move into self-employment. This income pattern offers us a potential clue to understand why foreign STEM graduates are less likely to be self-employed.

This paper provides substantial contributions to the literature and implications to the U.S. immigration policies. Firstly, this is the first paper considering the self-employment differential across college majors for aliens to our knowledge. Only if the facts are known, the direction of immigration reform or policy adjustment could be clearer. For example, considerable literature shows that foreign STEM graduates bring higher productivity and more innovation (Winters 2014; Peri et al. 2015); thus pro-STEM immigration policies could be established based on those studies[26]. Given that entrepreneurial activities enhance economies, if we know how college majors are

[26] Foreign STEM graduates have 29-month extension for optional practical training (OPT), while non-STEM college graduates only have 12-moth OPT extension at most.

correlated with self-employment rates, some new insights could be provided to the policymakers, and hence we could obtain more benefits from educated immigrants.

Second, this paper implies that foreign STEM workers could increase productivity and innovation, but which does not mean that they are more likely to be involved in entrepreneurial activities. This finding introduces a trade-off of reducing the immigration barrier for foreign STEM graduates. On one hand, importing more STEM human resources increases productivity and innovation. On the other hand, having more STEM immigrants may lower the self-employment rates. Therefore, it suggests policymakers should consider both advantages and potential caveats of STEM immigrants. If some other majors have higher self-employment rates, policies need lean towards these fields to maximize the benefits. An even more complex policy consideration is the combination of the advantages from STEM and the majors with high self-employment rates in some ways. For example, providing subsidies and/or reducing immigration barriers for those startups co-founded by graduates educated in STEM and fields with high self-employment rates.

Lastly, income differentials tell us why foreign STEM graduates are less likely to be self-employed from one aspect. If foreign STEM workers care more about their current benefits, it makes sense that they are less likely to be self-employed since they can earn more from employment at present, although self-employment may bring higher returns in the future. Besides, there could be a lot of other possibilities making foreign STEM less likely to be self-employed, which needs to be examined in the future studies.

The next section is the literature review, which provides some conceptual background. Section 3 describes our empirical approach and data. Section 4 provides empirical results and discussions. Section 5 concludes.

2 Literature Review and Conceptual Background

Aliens are more likely to be self-employed than their native counterparts in the U.S. (Borjas 1986). Figure 2.1 shows the mean self-employment rates for aliens and natives adjusted by sampling weights to ensure national representativeness. The sample is restricted to workers aged from 25 to 61 residing in the U.S. in 2013. 11.3% aliens are self-employed, while 9% for natives. The substantial difference indicates that the self-employment differential between foreigners and natives does not vanish over time. In order to explain this long-lasting difference, plenty of studies have identified determinants of being self-employed, including educational attainments, original nationalities, demographic characteristics, and utility heterogeneity. These determinants will be discussed in the literature review as follows.

Some researchers believe that education is one of the most important determinants for self-employment (Aronson 1991; Fairlie and Meyer 1996; Lofstrom 2004). Self-employment differentials exist by education level for foreign workers. Figure 2.2 presents the differences in mean self-employment rates for foreign and native workers aged from 25 to 61 by their educational attainments. The dark bars indicate the mean self-employment rates for foreigners with at least a college degree and without a college degree. The light bars show the mean percentages of self-employment for native workers with at least a college degree and without a college degree. Native workers present not much difference by the educational categories with similar rates around 9% as in Figure 2.1. However, self-employment differentials by educational groups are substantial for foreign workers compared to natives. The self-employment rate for foreigners without a college degree is 11.9%, while only 10.1% of foreigners with at least a college degree are self-employed.

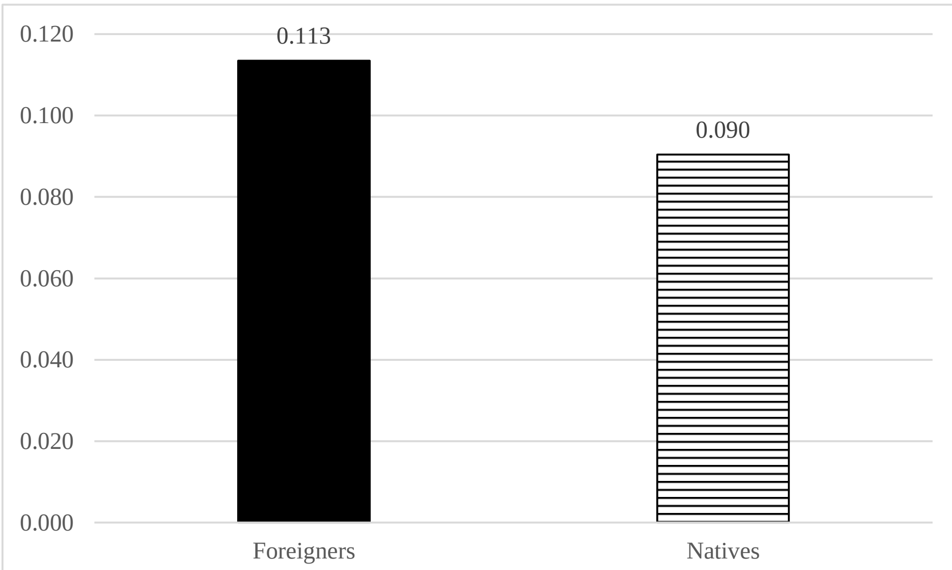

Figure 0.1: Self-employment rates by foreign status. Source: Author.

Self-employment rates for foreign workers vary across their original nationalities through different ways (Fairlie and Meyer, 1996; Lofstrom 2004). Self-employment rates are naturally higher in some source countries. Immigrants from these countries may also more likely be self-employed in the host countries (Yuengert 1995; Lassmann and Busch 2015). The positive association mainly comes from the experience and traditions of commerce brought from their home countries (Light 1984; Portes and Zhou 1991; Meyer 1990; Linan and Fernandez-Serrano 2014). However, Fairlie and Meyer (1996) empirically find that this correlation is insignificant, and argue that Yuengert (1995) overstates the correlation due to econometrics issue. Light (1972, 1979), Sowell (1981), and Moore (1983) argue that variations in disadvantages over source countries influence self-employment rates, including language barriers, discrimination, etc. Lofstrom (2004) finds that the self-employed immigrants have better English ability than their salaried counterparts. Besides, discrimination against immigrants from some specific countries in the labor market could increase self-employment rates of that national group (Lofstrom 2004; Constant and Zimmerman 2006; Fairlie 2006). Lofstrom (2004) also argues that the education estimates will be biased if countries of origin are not included in the self-employment decision regression. Therefore, original nationalities and English proficiency are always included in our empirical analysis.

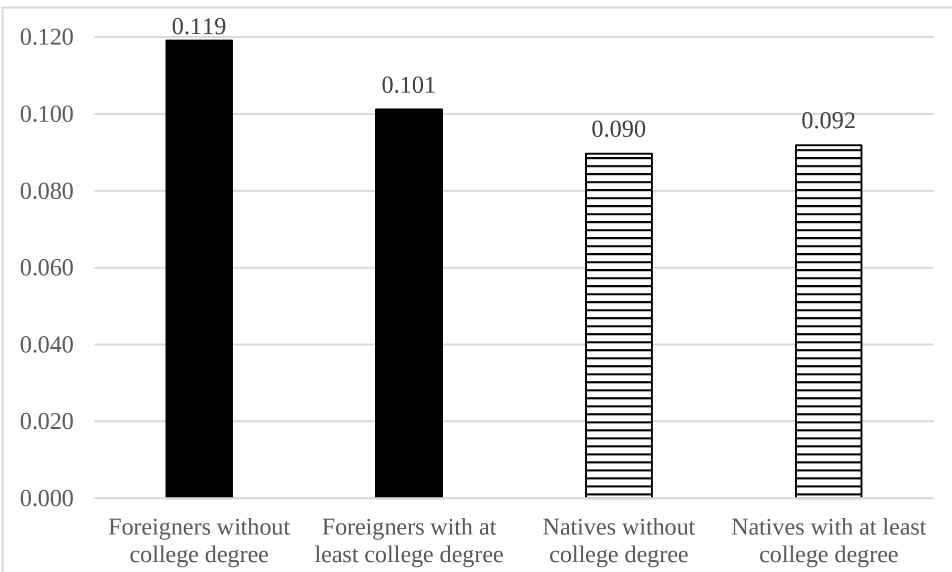

Figure 0.2: Self-employment rates by foreign status and education. Source: Author.

Figure 2.3 presents the self-employment rates for foreigners residing in the U.S. over their source country groups. Korea, Eastern Europe, and Canada have the highest rates. Aliens from Philippines have the lowest rates of self-employment, which is very close to the previous literature.[27]

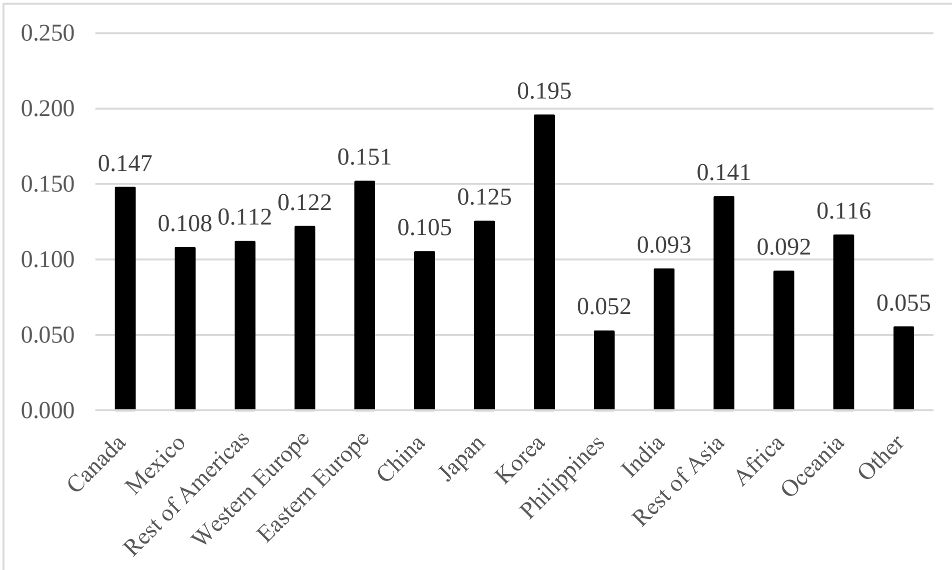

Figure 0.3: Self-employment rates by original country groups. Source: Author.

Demographic and socioeconomic attributes are the most used controls included in self-employment decision research (Lofstrom 2004; Fairlie and Meyer 1996; Blanchflower 2000; Djankov et al. 2005; 2006). Specifically, Devine (1994) finds that females are less likely to be self-employed than males, but females increasingly have become involved in small businesses over time. Some researchers claim that using ethnic and racial variables may bias the estimates and lose information (Lieberson and Waters 1988; Fairlie and Meyer 1996), which is also a reason why we only include countries of origin in our analysis.

Fields of study in college are overlooked in most self-employment research to our knowledge. Existing literature only shows that college major choices are strongly

[27] See Fairlie and Meyer (1996) as an example.

associated with earnings (Arcidiacono 2004; Winters and Xu 2014; Eide et al. 2015). However, many studies on self-employment build their theories on earning differentials. Kihlstrom and Laffont (1979) construct a general equilibrium model of firm formation. They model the self-employment decision by comparing the uncertain profits of entrepreneurship with the wage determined in the competitive labor market. Yuengert (1995) builds a self-employment choice model on Box-Cox transformations of earnings. Fairlie and Meyer (1996) finds that self-employment differentials across ethnic/racial groups are positively correlated with the earnings differences between the self-employed and the salaried. Portes and Zhou (1996) discuss the debate on self-employment returns, and argue that the earnings difference between the self-employed and the salaried are related to the choice of functional form of the earnings equations. They argue that the log-linear form fits the data better, but wastes the information of outliers.

Therefore, college major fields are likely associated with the self-employment decision through earnings. This paper focuses on self-employment of foreign STEM graduates, since they are productive and innovative, and usually are reported as high income receivers, drawing attention of policymakers (Winters 2014). However, the nature of innovation does not mean foreign STEM graduates are more likely to be self-employed. Figure 2.4 shows the mean self-employment rates for STEM and non-STEM graduates separately for foreigners and natives aged from 25 to 61. The native group still shows not much difference between the two categories as in the previous comparisons. The foreign group provides a surprising result that STEM graduates are substantially less self-employed nationwide with a rate of only 8.7%, even less than the native average. Foreign non-STEM on the other side presents a significantly high rate of self-employment. Thus, the differential in self-employment between STEM and non-STEM graduates only exists among immigrants.

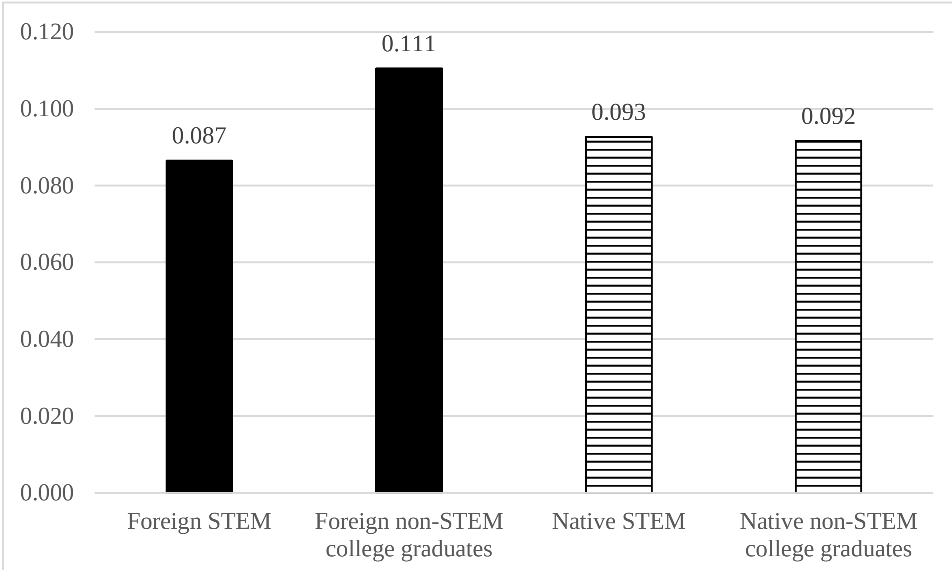

Figure 0.4: Self-employment rates by foreign status and STEM status. Source: Author.

Following the existing literature, we try to use potential earnings differentials to explain the self-employment decisions of foreign STEM and non-STEM graduates, and provide some theoretical thoughts as follows. Considering a static framework, if the potential earnings difference between self-employment and employment is quite substantial, this opportunity cost may prevent a worker from entering the type of work with lower income. Empirical studies have a long-lasting debate on whether self-employment has higher return. The debate usually depends on the functional form of the earnings equations (Portes and Zhou 1996) and the measure of self-employment (Levine and Rubinstein 2013). Leaving this debate behind, no matter whether the current incomes for the self-employed are higher or lower, high expected returns would be the major reason for people still wanting to become self-employed. However, the shape of the utility function is neglected by most studies, which could play a role as well. Considering risk preferences and intertemporal preferences, risk averse people are less likely to be self-employed since there exists much uncertainty in the future (Kihlstrom and Laffont 1979).

According to some empirical studies on risk preference, females are significantly more risk averse than males (Jianakoplos and Bernasek 1998). This is a good explanation for why females are less likely to be self-employed than males.[28] However, Halek and Eisenhauer (2001) find that the self-employed are not significantly different from their salaried counterparts facing speculative risks, which implies that risk averse people do not sort into employment rather than self-employment, thus the employed and self-employed averagely have the same attitude to the future income. Therefore, the current earnings differential still plays a larger role than the risk preference in the self-employment decision.

Institutional constraints are also correlated with self-employment rates, but it cannot explain the low rate of foreign STEM graduates. STEM and non-STEM immigrants with college degrees are facing the similar immigration policy. Theoretically, OPT extension, H-1B, E-B5, L1, and even F1 visas could be used to run a startup in order to obtain the permanent residency in the U.S. STEM graduates are even preferentially treated since they have a longer OPT extension. For those who want to run their own business, it costs less for STEM graduates to be self-employed since they have longer time of being legal in the U.S. Thus, STEM graduates should be more likely to be self-employed, considering the institutional constraints. However, Figure 2.4 tells us that the

[28] Fairlie and Meyer (1996) finds that female self-employment rates are usually about 55% of male rates.

current immigration policy does not make STEM graduates more entrepreneurial. Therefore, to sum up, current earning differentials are more reasonable to explain the variations in self-employment rates across college majors.

3 Empirical Framework and Data

This paper uses the linear probability model (LPM) to fit self-employment differentials across different college majors. The main model is:

$$Pr(SE_i = 1|\boldsymbol{m}_i, \boldsymbol{X}_i) = \boldsymbol{m}_i'\boldsymbol{\alpha} + \boldsymbol{X}_i\boldsymbol{\beta} \tag{6}$$

where the subscript *i* denotes individual observations. $\boldsymbol{m}_i$ is a vector of dummies for college majors, and $\boldsymbol{X}_i$ is a matrix of individual characteristics. The dependent variable SE_i is a dummy variable indicating whether a worker is self-employed or employed. In order to have a general to specific view of the self-employment differentials across different majors, three setups are assigned for the vector $\boldsymbol{m}_i$. First, it only includes the indicator of STEM college graduates, in order to compare the difference of self-employment between STEM college graduates and the other major graduates. The second setup for $\boldsymbol{m}_i$ is a vector of dummies for seven major graduate groups, including STEM, business, education, health, liberal arts, and social science graduates. The omitted group is college graduates in all other fields. Lastly, the largest 44 detailed majors are represented by $\boldsymbol{m}_i$.

The coefficients of our interest are $\boldsymbol{\alpha}$. For the first setup of $\boldsymbol{m}_i$, it indicates the self-employment differentials between STEM graduates and non-STEM graduates. For the second setup of $\boldsymbol{m}_i$, $\boldsymbol{\alpha}$ shows the average differences between different major groups and the base group. For the last setup, $\boldsymbol{\alpha}$ shows the average differences between detailed majors and the base category. Then it is reasonable to use these estimates to rank the marginal probability of being self-employed across the detailed majors.

One major concern of this model is that industries could be correlated with self-employment rates, and also associated with college majors. Thus, the second specification of our model controls for the industry fixed effects. Moreover, considering local economic variation, Metropolitan Statistical Area (MSA) fixed effects are also included in the model, which is:

$$Pr(SE_{idc} = 1|\boldsymbol{m}_{idc}, \boldsymbol{X}_i) = \boldsymbol{m}'_{idc}\boldsymbol{\alpha} + \boldsymbol{X}_{idc}\boldsymbol{\beta} + \tau_d + \pi_c \tag{7}$$

where subscript *d* indexes industries, and *c* denotes MSAs and non-MSAs. τ_d is industry fixed effects, and π_c is MSA fixed effects.

As discussed in the previous sections, earnings differentials may play a role in the self-employment decision for foreign STEM graduates. To have a formal test on the income differences across majors, the model below is proposed:

$$\begin{aligned} ln(income_{idc}) \\ &= \theta SE_{idc} + \delta STEM_{idc} + \rho SE_{idc} \times STEM_{idc} + \boldsymbol{X}_{idc}\boldsymbol{\beta} \\ &+ \tau_d + \pi_c + \epsilon_{idc} \end{aligned} \tag{8}$$

where the dependent variable is the log of total personal earned income in the previous year. $STEM_{idc}$ is a dummy for STEM college graduates. It will be extended to more detailed majors to compare the earnings differentials among different majors. ρ is the coefficient of the most interest, which captures the average income difference between self-employed STEM workers and other workers. If this coefficient is negative, it means there is a drop in earnings if STEM graduates choose self-employment, which could explain the low self-employment rate for foreign workers with college degree in STEM as discussed in the previous section,

This paper uses the 1% IPUMS (Ruggles et al., 2010) 2013 ACS sample. Our sample only includes the foreign-born[29] employed and self-employed workers aged from 25 to 61 with at least college degree. All empirical steps are adjusted by personal sampling weights (PERWT in IPUMS) to ensure the results are representative.

The class of worker (CLASSWKR) variable is employed to define the self-employed. The detailed version of this variable (CLASSWKRD) is also used to identify those who are incorporated self-employed, since it may capture more entrepreneurial spirit (Levine and Rubinstein 2013). The workers either in MSA or non-MSA areas are included in the sample by recoding the MSA identifier with the state variable. The last year earned income (INCEARN) is used as the income measure.

The STEM graduates are defined mainly based on the STEM list from the U.S. Immigration and Customs Enforcement[30]. However, this list is not identical with our

[29] We define foreign-born people using birthplace (BPL) variable from the IPUMS. If BPL > 56, people are treated as foreigners, including who are born in the US outlying areas or territories (American Samoa, Guam, Puerto Rico, U.S. Virgin Islands).
[30] The list is available at http://www.ice.gov/doclib/sevis/pdf/stem-list.pdf

definition because of the major coding difference with the ACS[31]. In our STEM definition, we just consider the majors in the first degree, which is consistent with our definition of the other major groups. ACS also provides the information on people's majors in their second degrees. This would make it possible that we could have alternative definitions of STEM. A narrower and a broader version of STEM will be used as sensitivity analysis. The narrower version is just considering those with both degrees in STEM and non-STEM. The broader version will also treat the observations only with a second major in STEM as STEM graduates.

There are also six broad major groups defined, including business, education, health, liberal arts, social science, and other majors. Business group includes non-STEM majors with fields of degree in Business. Education group includes non-STEM majors with fields of degree in Education Administration and Teaching. Health group includes non-STEM majors with fields of degree in Medical, Health Sciences and Services. Liberal arts group includes non-STEM majors with fields of degree in Area, Ethnic, and Civilization Studies, Communications, Linguistics and Foreign Languages, English Language, Literature, and Composition, Liberal Arts and Humanities, Library Science, Interdisciplinary and Multi-Disciplinary Studies (General), Philosophy and Religious Studies, Theology and Religious Vocations, Fine Arts, History. Social science group includes non-STEM majors with fields of degree in Law, Psychology, Criminal Justice and Fire Protection, Public Affairs, Policy, and Social Work, Social Sciences. Other major group includes non-STEM majors with all fields of degree not mentioned.

[31] The list of majors in STEM of our version is provided in the Appendix Table 1.

A standard set of demographic characteristics is controlled for in all regressions, including a quadratic polynomial of age, dummies for female, married, an interaction of female and married, naturalized citizen, children present in household, and three dummies for educational attainments. Four dummies for English ability are also included since it seems to be naturally correlated to the self-employment decision. Additionally, 13 dummies of original nationality groups are included as well. They are Canada, Mexico, Rest of Americas, Western Europe, Eastern Europe, China, Japan, Korea, Philippines, India, Rest of Asia, Africa, and Oceania.[32] Observations from regions other than above are omitted as a group of other. Four years-in-the-USA intervals are also included. Ethnicity and race are excluded from our analysis since they are potentially highly correlated to the original nationalities, and for the reasons discussed in the previous section.

In order to control for the industry fixed effects, Ind1990 from IPUMS are used to identify industries, which is a three-digit identifier. There are 223 industries in our sample. Table 2.1 provides the summary statistics for foreign-born STEM and non-STEM college graduates.

[32] These dummies are created based on the recoded birthplace (BPL) variable.

Table 2.1: Summary Statistics for Analytical Sample

	STEM		Non-STEM	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
<u>A. Outcome Variables</u>				
Self-Employed	0.087	0.281	0.111	0.314
Incorporated Self-Employed	0.047	0.211	0.050	0.217
Log Annual Earnings	11.037	1.036	10.581	1.117
<u>B. Individual Control Variables</u>				
Age	41.405	9.683	42.105	9.845
Bachelor's Degree	0.461	0.498	0.651	0.477
Master's Degree	0.334	0.472	0.252	0.434
Professional Degree	0.082	0.275	0.062	0.241
Doctoral Degree	0.122	0.328	0.036	0.185
Female	0.313	0.464	0.600	0.490
Married	0.750	0.433	0.667	0.471
Does Not Speak English	0.004	0.065	0.010	0.101
Speaks only English	0.206	0.405	0.274	0.446
Speaks English very well	0.586	0.493	0.508	0.500
Speaks English well	0.170	0.376	0.158	0.364
Speaks English not well	0.033	0.179	0.050	0.218
Canada	0.027	0.163	0.036	0.187
Mexico	0.040	0.196	0.070	0.255
Rest of Americas	0.126	0.332	0.223	0.416
Western Europe	0.078	0.268	0.101	0.301
Eastern Europe	0.082	0.275	0.086	0.280
China	0.123	0.328	0.071	0.257
Japan	0.012	0.110	0.022	0.147
Korea	0.035	0.183	0.047	0.213
Philippines	0.042	0.202	0.091	0.288
India	0.255	0.436	0.091	0.287
Rest of Asia	0.119	0.324	0.090	0.286
Africa	0.055	0.228	0.064	0.246
Oceania	0.005	0.070	0.007	0.083
Other	0.000	0.016	0.001	0.023
Naturalized citizen	0.508	0.500	0.545	0.498
Has Children	0.553	0.497	0.527	0.499
Years in USA 0-5	0.176	0.381	0.127	0.333
Years in USA 6-10	0.146	0.353	0.124	0.329
Years in USA 11-15	0.176	0.381	0.156	0.363
Years in USA 16-20	0.129	0.335	0.121	0.327
Years in USA 20+	0.373	0.484	0.471	0.499
<u>C. Non-STEM Broad Major Categories</u>				
Business Graduate			0.335	0.472
Education Graduate			0.093	0.290
Health Graduate			0.116	0.320
Liberal Arts Graduate			0.212	0.409
Social Science Graduate			0.198	0.399
Other Major Graduate			0.046	0.210
Observations	30,620		46,681	

Note: Analytical sample includes foreign-born college graduates ages 25-61 who are employed or self-employed. All results are adjusted by personal weight.

Panel A shows the raw means and standard deviations of the outcome variables. As shown in Figure 2.1, non-STEM foreign-born graduates are more likely to be self-employed than their STEM counterparts. The raw difference is about 0.024, which is statistically significant. The difference of incorporated self-employed rates between STEM and non-STEM is less substantial. Moreover, STEM workers earn significantly more than non-STEM workers, but one could not conclude an association between self-employed rates and annual earnings based on the raw means since the raw mean of annual earnings contains the information for both the employed and the self-employed. More formal tests will be conducted to study the correlation in the following section.

Panel B provides the summary statistics for the individual characteristics controls. According to the demographic attributes, STEM graduates are younger than their non-STEM counterparts on average. The educational attainments characterize an unsurprising fact that Bachelor’s degrees are concentrated in non-STEM fields, while STEM graduates obtain more degrees of higher levels, showing that people in STEM fields are more academically oriented. More native English speakers sort into non-STEM fields, but STEM English learners generally have better English skills than their non-STEM counterparts. This is somewhat unexpected at first glance since a lot of non-STEM fields seems to require higher communication skills than STEM fields, but considering the intra-field competition, non-native speakers may have more advantages to compete if they have better English skills in STEM fields which is less pursued by native speakers. China and India are highlighted in the composition of original nationalities for STEM and non-STEM fields. These are the only two countries with a higher proportion in STEM than in non-STEM fields. Panel C shows the composition of non-STEM broad major categories. All the summary statistics are adjusted by personal sampling weights to ensure they are nationally representative.

4 Empirical Results

4.1 Self-employment Differentials between Foreign STEM and Non-STEM fields

Models (1) and (2) are estimated by OLS to investigate whether there is a real difference of self-employment rates across majors. First, we only consider the difference between STEM and non-STEM workers. Table 2.2 presents the self-employment differentials between STEM and non-STEM college graduates for three specifications.[33]

Columns (1) and (4) show the estimates for our baseline model without any fixed effect controls. The dependent variables are dummies for self-employed and incorporated self-employed respectively. The estimates are highly significant and large in magnitudes. One advantage of LPM is that the estimates could be directly interpreted as the marginal effects. Thus, Column (1) indicates that the average probability of being self-employed for STEM graduates is less than that for non-STEM graduates by 3.06 percent, *ceteris paribus*. After controlling for individual characteristics, the self-employed differential becomes even larger than the raw difference with a value of 2.4 percent. Surprisingly, the incorporated self-employment differential becomes highly significant and larger in magnitude after controlling for personal attributes. Compared with the raw difference of

[33] The estimates of detailed controls are presented in the Appendix Table 2.

0.3 percent, Column (4) indicates that STEM graduates are less likely to be incorporated self-employed by 1.37 percent than non-STEM graduates, *ceteris paribus*.

Table 2.2: Self-Employment Differentials Between STEM and All Non-STEM College Graduates

Outcome:	Self-Employed			Incorporated Self-Employed		
	(1)	(2)	(3)	(4)	(5)	(6)
STEM Graduate	-0.0306***	-0.0160***	-0.0153***	-0.0137***	-0.0091***	-0.0087***
	(0.0039)	(0.0034)	(0.0033)	(0.0021)	(0.0023)	(0.0022)
Industry Fixed Effects	No	Yes	Yes	No	Yes	Yes
MSA Fixed Effects	No	No	Yes	No	No	Yes

Notes: The omitted base group includes college graduates with bachelor's degrees in all non-STEM fields. All regressions include individual controls listed in Table 2.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by MSA.

Industry fixed effects are added to the baseline model, and Columns (2) and (5) of Table 2.2 provide the results. The estimated differentials are still highly significant and large, but the magnitudes of absolute values decrease by about 50% and 34% for being self-employed and incorporated self-employed, respectively. One possibility is because of the nature of different industries that some industries naturally have high self-employed rates, while it is not apt to be self-employed in some other industries.[34] Besides, one cannot rule out that majors, self-employment, and industries are jointly determined, which could bias our estimates. Columns (3) and (6) report the estimated differentials also controlling for MSA fixed effects. For both self-employment measures, the estimates are still nontrivial with just a little drop in the magnitudes. Even though similar estimates are obtained from the second and third specifications, we prefer the latter with more observable controls.

[34] Industry ranking of self-employment rates are presented in Appendix Table 3.

4.2 Self-employment Differentials across Broad Major Groups

The foreign STEM graduates have substantially lower self-employment rates, who are well documented as being innovative and productive. It would raise a question on what major groups have higher self-employment rates? Or whether a specific major group has an extremely high rate of self-employment, pushing up the average of the non-STEM category? Table 2.3 answers the questions by providing the results of self-employment differentials for broad major categories. The preferred specification is used, controlling for industry fixed effects and MSA fixed effects. All majors are grouped into 7 categories, including STEM, business, education, health, liberal arts, social science, and other major graduates as the omitted reference group. For the probability of being self-employed, all major groups are less likely to be self-employed than the base group, but liberal arts, social science, and health groups do not significantly have less probability of being self-employed than the reference group, and they have the highest three self-employment rates among the non-reference major groups. For the groups with lowest self-employment rates, the STEM group is highlighted with the lowest probability of being self-employed, following by education and business graduate groups.

For the probabilities of being incorporated self-employed, the picture does not change much, but the order shuffles a little. Following the reference group, from high to low probabilities of being incorporated self-employed, they are health, social science, business, liberal arts, STEM, and education graduates. On one hand, this might be due to the nature of major groups. For example, the health field naturally has more incorporated entities to protect the rapidly increasing biomedical patents and new techniques. On the other hand, as discussed earlier, incorporated self-employment may measure different aspects from self-employment. Incorporated self-employment may capture more entrepreneurial aspects. However, STEM graduates still rank behind four broad major groups, namely health, social science, business, and liberal arts. Therefore, controlling for exhaustive covariates and using an alternative measure of self-employment, STEM graduates are always correlated with low rates of self-employment.

Table 2.3: Self-Employment Differentials for Broad Major Categories

Outcome:	Self-Employed	Incorporated Self-Employed
STEM Graduate	-0.0271***	-0.0154**
	(0.0082)	(0.0062)
Business Graduate	-0.0191**	-0.0054
	(0.0092)	(0.0068)
Education Graduate	-0.0256***	-0.0186**
	(0.0082)	(0.0074)
Health Graduate	-0.0108	-0.0026
	(0.0070)	(0.0063)
Liberal Arts Graduate	-0.0029	-0.0106
	(0.0083)	(0.0068)
Social Science Graduate	-0.0071	-0.0050
	(0.0075)	(0.0063)
Industry Fixed Effects	Yes	Yes
MSA Fixed Effects	Yes	Yes

Notes: The omitted base group includes college graduates with bachelor's degrees in all other non-listed fields. All regressions include individual controls provided upon request. Standard errors in parentheses are robust to heteroskedasticity and clustered by MSA.

4.3 Self-employment Differentials for Selected Detailed Majors

Given the picture that STEM graduates are less likely to be self-employed, it would be interesting to look at the variation among detailed majors. In order to have a clearer presentation, Table 2.4 ranks the likelihoods of being self-employed and incorporated self-employed respectively based on the magnitudes of the estimated coefficients of model (2). 45 detailed majors with largest numbers of observations are selected, which cover 80.66% of the whole sample. Including the remaining small majors would give noisy estimates. To have a consistent reference group, the omitted category is the same as it in Table 2.3. However, architecture is one of the majors included in the reference category in Table 2.3, but it is also one of the 45 biggest majors. Thus, we drop it from the rankings, and leave it in the reference group as we did in Table 2.3 to maintain consistency. Eventually, we have 44 largest majors in the rankings.

The ranking for self-employment presents a similar pattern we get from the previous results. The estimates of the beginning and the end in the ranking are significant. The majors in the intermediate part of the list are not statistically different. However, only looking at the ranks, only three STEM majors enter the top 20, which are Pharmacy, Pharmaceutical Sciences, and Administration (6^{th}), Industrial and Manufacturing Engineering (19^{th}), and Biology (20^{th}). Most STEM majors stay at the end of the ranking, and the differentials are substantial.

The ranks of STEM majors improve a little bit in incorporated self-employment (the average rank of STEM majors in self-employment is 31, and 28.06 in incorporated self-employment), but it does not qualitatively change the trend that STEM graduates are less likely to sort into incorporated self-employment (the correlation of ranks between self-employment and incorporated self-employment is about 0.68). More STEM majors enter the top 20 in incorporated self-employment, including Pharmacy, Pharmaceutical Sciences, and Administration (2^{nd}), Chemical Engineering (9^{th}), Biology (12^{th}), Chemistry (14^{th}), and Biochemical Sciences (15^{th}). However, STEM majors still occupy the lower end of the ranking with significant differentials and relatively large magnitudes.

Table 2.4: Self-Employment Differentials for Selected Detailed Major Categories

Outcome:	Self-Employed			Incorporated Self-Employed		
	Rank	Coefficient	Std. Error	Rank	Coefficient	Std. Error
Music	1	0.0687***	(0.0183)	7	-0.0031	(0.0150)
Treatment Therapy Professions	2	0.0369**	(0.0183)	1	0.0381**	(0.0148)
Commercial Art and Graphic Design	3	0.0319**	(0.0145)	21	-0.0107	(0.0082)
Fine Arts	4	0.0245	(0.0229)	5	-0.0004	(0.0172)
Journalism	5	0.0123	(0.0169)	37	-0.0204*	(0.0117)
Pharmacy, Pharmaceutical Sciences, and Administration	6	0.0074	(0.0153)	2	0.0159	(0.0131)
Psychology	7	0.0028	(0.0090)	13	-0.0069	(0.0080)
Philosophy and Religious Studies	8	-0.0012	(0.0266)	43	-0.0309***	(0.0102)
Multi-disciplinary or General Science	9	-0.0029	(0.0142)	19	-0.0105	(0.0096)
Theology and Religious Vocations	10	-0.0046	(0.0189)	22	-0.0116	(0.0177)
Economics	11	-0.0052	(0.0082)	4	0.0041	(0.0069)
Nursing	12	-0.0067	(0.0075)	8	-0.0038	(0.0063)
Political Science and Government	13	-0.0077	(0.0121)	11	-0.0049	(0.0081)
English Language and Literature	14	-0.0079	(0.0111)	6	-0.0028	(0.0088)
General Business	15	-0.0081	(0.0107)	3	0.0049	(0.0082)
Hospitality Management	16	-0.0085	(0.0197)	17	-0.0085	(0.0156)
French, German, Latin and Other Common Foreign Language Studies	17	-0.0104	(0.0135)	16	-0.0084	(0.0099)
General Education	18	-0.0145	(0.0098)	26	-0.0125	(0.0084)
Industrial and Manufacturing Engineering	19	-0.0166	(0.0150)	27	-0.013	(0.0121)
Biology	20	-0.017*	(0.0089)	12	-0.0057	(0.0074)
Business Management and Administration	21	-0.019*	(0.0098)	10	-0.0045	(0.0076)
Chemistry	22	-0.0191	(0.0143)	14	-0.0075	(0.0104)
Chemical Engineering	23	-0.0195	(0.0134)	9	-0.004	(0.0090)
Marketing and Marketing Research	24	-0.0196	(0.0175)	20	-0.0106	(0.0149)
Electrical Engineering Technology	25	-0.0198	(0.0168)	31	-0.0176	(0.0127)
Sociology	26	-0.0204	(0.0136)	28	-0.0148	(0.0102)
Finance	27	-0.0206**	(0.0102)	24	-0.0121*	(0.0072)
Biochemical Sciences	28	-0.0225	(0.0165)	15	-0.0078	(0.0122)
Communications	29	-0.0232	(0.0144)	23	-0.0118	(0.0099)
Criminal Justice and Fire Protection	30	-0.026	(0.0167)	29	-0.015*	(0.0082)
Social Work	31	-0.026*	(0.0152)	25	-0.0123	(0.0109)
Accounting	32	-0.0264***	(0.0096)	18	-0.0091	(0.0067)
General Engineering	33	-0.0268**	(0.0128)	42	-0.0261***	(0.0070)
Elementary Education	34	-0.0318***	(0.0107)	34	-0.0196**	(0.0078)
History	35	-0.0333**	(0.0146)	36	-0.0198***	(0.0075)
Mechanical Engineering	36	-0.0336***	(0.0095)	38	-0.0204***	(0.0077)
Electrical Engineering	37	-0.0371***	(0.0085)	41	-0.0238***	(0.0077)
Computer and Information Systems	38	-0.0374***	(0.0108)	32	-0.0181**	(0.0088)
Mathematics	39	-0.0381***	(0.0126)	30	-0.0164	(0.0103)
Liberal Arts	40	-0.0388***	(0.0115)	35	-0.0197**	(0.0099)
Computer Engineering	41	-0.0389***	(0.0126)	33	-0.0189**	(0.0090)
Computer Science	42	-0.0409***	(0.0090)	39	-0.0208***	(0.0071)
Physics	43	-0.0573***	(0.0105)	44	-0.0339***	(0.0082)
Civil Engineering	44	-0.0684***	(0.0100)	40	-0.0228**	(0.0090)
Industry Fixed Effects		Yes			Yes	
MSA Fixed Effects		Yes			Yes	

Notes: The omitted base group includes college graduates with bachelor's degrees in fields outside of STEM, Business, Education, Health, Liberal Arts, and Social Science, i.e., all other fields not listed in Table 2.3; controls for smaller detailed majors in the main categories from Table 2.3 are included but not reported. Regressions include individual controls provided upon request. Standard errors in parentheses are robust to heteroskedasticity and clustered by MSA. STEM majors are bold.

4.4 Annual Earning Differentials between STEM and Non-STEM graduates

As discussed in Section 2, we try to explain the correlation between major decision and self-employment decision through earning differentials. Table 2.5 presents the estimates for Model (3). Similar as in Table 2.2, Table 2.5 contains three specifications. Columns

(1) and (4) do not include industry and MSA fixed effects. Columns (2) and (5) add industry fixed effects, and Columns (3) and (6) includes all fixed effects. The dependent variable is log personal annual income. Columns (1) to (3) are self-employment specifications, and Columns (4) to (6) are incorporated specifications.

All level effects are significant for self-employment specifications. The interaction effect in Column (1) is significant as well. Being self-employed is correlated with 0.26 less log points of annual income for non-STEM graduates, *ceteris paribus*. STEM salaried workers have 0.26 more log points of annual income than non-STEM employees, *ceteris paribus*. However, the interaction of STEM and self-employed is negative, showing that incomes for STEM workers will drop further if they become self-employed.

Column (2) controls for industry fixed effects, and Column (3) controls for all fixed effects. Columns (2) and (3) present similar estimates. We still choose the latter as our preferred specification as we did in the previous tables. The interaction estimates become insignificant after controlling for the fixed effects, and the magnitudes are cut by half. Nonetheless, the level effects of self-employment and STEM remain same sign respectively, and similar in magnitudes. Thus, if without the interaction term, the net effect on income for STEM graduates of being self-employed is around zero, which implies that STEM graduates will earn much less if they choose to be self-employed.

The incorporated specifications tell the similar story with a little variation. The level effects and interaction effects are significant for all incorporated specifications. Columns (4), (5) and (6) show that the annual income is significantly higher of being incorporated self-employed for non-STEM graduates, *ceteris paribus*. However, this earnings advantage will significantly decease for STEM graduates. For non-incorporated, STEM graduates earn significantly more than their non-STEM counterparts, *ceteris paribus*. Nevertheless, this earning advantage will also substantially drop if STEM graduates choose to be incorporated self-employed.

Table 2.5: Log Annual Earnings Differentials Using STEM and Non-STEM Categories

	(1)	(2)	(3)		(4)	(5)	(6)
SE	-0.2648***	-0.1435***	-0.1385***	Incorp	0.0824**	0.1882***	0.1940***
	(0.0388)	(0.0295)	(0.0291)		(0.0365)	(0.0323)	(0.0307)
STEM	0.2639***	0.1332***	0.1341***	STEM	0.2723***	0.1379***	0.1387***
	(0.0208)	(0.0114)	(0.0098)		(0.0198)	(0.0107)	(0.0093)
STEM × SE	-0.1131**	-0.0525	-0.0478	STEM × Incorp	-0.1543***	-0.0910**	-0.0834**
	(0.0449)	(0.0438)	(0.0437)		(0.0420)	(0.0357)	(0.0357)
Industry Fixed Effects	No	Yes	Yes		No	Yes	Yes
MSA Fixed Effects	No	No	Yes		No	No	Yes

Notes: The omitted group includes all college graduates with bachelor's degrees in non-STEM fields. Regressions include individual controls provided upon request. Standard errors in parentheses are robust to heteroskedasticity and clustered by MSA. SE denotes self-employment, and Incorp stands for incorporated self-employment.

Therefore, all the results tell us the same story that if STEM graduates sort into self-employment, their earnings will significantly drop. This could be a strong reason preventing foreign STEM graduates from being self-employed. Although the return of self-employment could be very high in the future, people with no preference on uncertainty will definitely enjoy higher current benefits. As noted in Section 2, some evidence shows that the employed and self-employed workers have similar risk preferences, thus self-employment decision depends on current incomes rather than expected incomes.

Table 2.6: Log Annual Earnings Differentials Using Selected Detailed Major Categories

	Self-employed		Incorporated	
	Level	Interaction	Level	Interaction
Accounting	0.0651*	0.1443	0.0883**	0.0594
	(0.0341)	(0.1183)	(0.0343)	(0.1377)
Biochemical Sciences	0.0596	0.2695	0.0764	0.2004
	(0.0547)	(0.2344)	(0.0551)	(0.2571)
Biology	0.1074***	0.1598	0.1285***	-0.0053
	(0.0375)	(0.1061)	(0.0375)	(0.1534)
Business Management and Administration	0.0460	-0.0187	0.0534	-0.0362
	(0.0326)	(0.1244)	(0.0336)	(0.1439)
Chemical Engineering	0.2797***	-0.3038*	0.2830***	-0.4558**
	(0.0432)	(0.1598)	(0.0408)	(0.1954)
Chemistry	0.0910**	-0.0968	0.1055***	-0.3087
	(0.0354)	(0.2445)	(0.0356)	(0.3571)
Civil Engineering	0.1834***	-0.0800	0.2043***	-0.1740
	(0.0317)	(0.1234)	(0.0340)	(0.1683)
Commercial Art and Graphic Design	0.0665	-0.3025	0.0214	-0.2936
	(0.0555)	(0.1861)	(0.0531)	(0.2260)
Communications	0.0728*	0.0601	0.0856**	-0.0154
	(0.0374)	(0.1886)	(0.0383)	(0.1824)
Computer and Information Systems	0.2108***	-0.0546	0.2125***	0.0931
	(0.0403)	(0.2151)	(0.0448)	(0.1809)
Computer Engineering	0.3477***	-0.0909	0.3611***	-0.2971*
	(0.0319)	(0.1660)	(0.0334)	(0.1729)
Computer Science	0.2488***	-0.2977*	0.2471***	-0.1092
	(0.0371)	(0.1667)	(0.0356)	(0.1668)
Criminal Justice and Fire Protection	-0.0457	-0.0616	-0.0278	-0.5941
	(0.0475)	(0.2711)	(0.0449)	(0.4287)
Economics	0.0886***	0.0153	0.0913***	0.0482
	(0.0278)	(0.1237)	(0.0313)	(0.1249)
Electrical Engineering	0.2602***	-0.0331	0.2724***	-0.0750
	(0.0315)	(0.1094)	(0.0322)	(0.1493)
Electrical Engineering Technology	0.2720***	0.1528	0.2836***	0.1780
	(0.0552)	(0.2568)	(0.0554)	(0.2486)
Elementary Education	0.0098	-0.6239	-0.0102	-0.3189
	(0.0475)	(0.4127)	(0.0562)	(0.2780)
English Language and Literature	-0.0227	-0.0990	-0.0316	-0.0438
	(0.0462)	(0.1483)	(0.0423)	(0.1951)
Finance	0.1368***	0.2258*	0.1592***	0.1540
	(0.0450)	(0.1287)	(0.0428)	(0.1325)
Fine Arts	-0.0703	-0.1751	-0.1046*	-0.1062
	(0.0533)	(0.2211)	(0.0626)	(0.1830)
French, German, Latin and Other Common Foreign Language Studies	0.0685	-0.2104	0.0566	-0.1845
	(0.0574)	(0.2692)	(0.0619)	(0.2265)
General Business	0.1252***	0.1194	0.1387***	0.0684
	(0.0300)	(0.1241)	(0.0298)	(0.1518)
General Education	0.0172	-0.1381	-0.0050	0.3507
	(0.0341)	(0.1729)	(0.0359)	(0.2171)
General Engineering	0.1728***	-0.0515	0.1918***	-0.2872**
	(0.0353)	(0.1392)	(0.0370)	(0.1333)
History	0.0061	-0.4671*	-0.0389	-0.0813
	(0.0466)	(0.2615)	(0.0539)	(0.2469)

Hospitality Management	-0.0217	-0.1767	-0.0072	-0.5716
	(0.0682)	(0.1829)	(0.0671)	(0.3532)
Industrial and Manufacturing Engineering	0.2782***	-0.1596	0.2740***	-0.0540
	(0.0459)	(0.1759)	(0.0439)	(0.2107)
Journalism	0.0086	-0.2049	-0.0208	0.0076
	(0.0716)	(0.2144)	(0.0675)	(0.1807)
Liberal Arts	0.0028	-0.1902	-0.0036	0.0756
	(0.0489)	(0.3097)	(0.0473)	(0.2502)
Marketing and Marketing Research	0.0627	-0.0325	0.0636	0.0841
	(0.0451)	(0.2439)	(0.0480)	(0.2590)
Mathematics	0.1456***	-0.0554	0.1559***	-0.0389
	(0.0441)	(0.1778)	(0.0449)	(0.1804)
Mechanical Engineering	0.2458***	-0.3331**	0.2334***	-0.1317
	(0.0367)	(0.1379)	(0.0405)	(0.1996)
Multi-disciplinary or General Science	0.1433***	0.1098	0.1700***	-0.1355
	(0.0446)	(0.2143)	(0.0469)	(0.1920)
Music	-0.1673**	0.3099**	-0.1229*	0.3364
	(0.0759)	(0.1427)	(0.0631)	(0.2636)
Nursing	0.2718***	0.1417	0.2890***	-0.0593
	(0.0443)	(0.1511)	(0.0446)	(0.1660)
Pharmacy, Pharmaceutical Sciences, and Administration	0.2217***	0.1408	0.2373***	-0.0710
	(0.0562)	(0.1890)	(0.0603)	(0.1759)
Philosophy and Religious Studies	-0.0943	0.1197	-0.0660	-0.0768
	(0.0591)	(0.2508)	(0.0615)	(0.5614)
Physics	0.1799***	-0.3175	0.1826***	-0.0702
	(0.0393)	(0.2397)	(0.0393)	(0.2164)
Political Science and Government	0.0119	0.1346	0.0250	0.1230
	(0.0453)	(0.1383)	(0.0438)	(0.1614)
Psychology	-0.0105	-0.1160	-0.0221	-0.1056
	(0.0298)	(0.1259)	(0.0312)	(0.1379)
Social Work	-0.0086	0.0024	0.0050	-0.0265
	(0.0744)	(0.2962)	(0.0702)	(0.3560)
Sociology	0.0055	-0.2184	-0.0062	0.0133
	(0.0510)	(0.2505)	(0.0487)	(0.3050)
Theology and Religious Vocations	-0.0814	0.0854	-0.0626	-0.0379
	(0.0703)	(0.2493)	(0.0692)	(0.2133)
Treatment Therapy Professions	0.2239***	0.2553	0.2313***	0.1803
	(0.0565)	(0.2081)	(0.0603)	(0.2079)
Self-Employment	-0.1300			
	(0.0938)			
Incorporated Self-Employment			0.1890	
			(0.1154)	
Industry Fixed Effects	Yes	Yes	Yes	Yes
MSA Fixed Effects	Yes	Yes	Yes	Yes

Notes: The omitted base group includes college graduates with bachelor's degrees in fields outside of STEM, Business, Education, Health, Liberal Arts, and Social Science, i.e., all other fields not listed in Table 2.3; controls for smaller detailed majors in the main categories from Table 2.3 are included but not reported. Regressions include individual controls provided upon request. Standard errors in parentheses are robust to heteroskedasticity and clustered by MSA.

4.5 Correlations between self-employment differentials and income differentials

The previous section provides a partial explanation for why foreign STEM graduates are less likely to be self-employed. To have more general and stronger support, we consider the correlations between self-employment differentials and income differentials for detailed college majors. If our point of view in the previous section is correct, then lower self-employment rates will be associated with higher incomes, *vice versa*. Thus, we need to test whether their correlation is negative or not.

Table 2.6 presents the estimates of log annual earnings differentials for our 44 selected detailed majors. Each major is interacted with being self-employed and incorporated self-employed respectively. The interpretations of the estimates are similar with those for Table 2.5. For example, chemical engineering graduates earn 0.28 log points more than the reference group if they are employed. However, if they change their mind and become their own bosses, their log annual income will drop more than 0.3 log points.

The correlations of the estimates from Table 2.4 and Table 2.6 are calculated. The correlation between self-employment differential estimates and log annual income level differential estimates is -0.402. This could be interpreted that the higher income for employees, the lower self-employment rates, *vice versa*. The correlation between self-employment differentials and log annual income interaction estimates is 0.36, which means the lower income for self-employed, the lower self-employment rates, vice versa. Therefore, these correlations nicely support our point in the last section that foreign college graduates are less likely to be self-employed since their earnings drop if they choose to be self-employed. Especially for STEM graduates, their very low rates of self-employment are associated with very high incomes for being employees, and also their low rates of self-employment are associated with significant decrease in incomes for being self-employed.

Correlations are also calculated for incorporated self-employment differentials and level effects and interaction effects of income differentials respectively. These correlations are much weaker than those for self-employment differentials, 0.020 for level differentials, and 0.213 for interaction estimates. Thus, there is not much correlation between the incorporated self-employment differentials and incomes of being employed, but lower incorporated self-employment rates are also associated with lower income of being incorporated self-employed.

4.6 Sensitivity Analysis

Table 2.7 presents the results for sensitivity analysis. The first three rows show the results for the alternative STEM definitions. The first row provides the replications of Column (3) in Table 2.2 and Column (3) in Table 2.5 as reference. In the second row of the broader STEM definition, STEM graduates are defined as the first degree is in STEM and/or the second degree is in STEM. The estimates do not change much. The third row provides the results for the stricter STEM definition, which only includes those who get their first and second degree both in STEM fields. The self-employment differential becomes smaller and trivial. This is because there are only 1,649 observations satisfying the stricter STEM definition. Compared with the preferred definition of 30,620 observations, and the broader definition of 31,158 observations, the small number of observations in stricter STEM specification makes the estimates are noisy. The fourth row shows the results for full time workers who usually work more than 35 hours per week. The results still do not change much.

Table 2.7: Sensitivity Analysis

Dependent Variable	Self-Employed	log (Income)		
Coefficients	STEM graduates	STEM graduates	SE	SE*STEM
STEM preferred definition	-0.0153***	0.1341***	-0.1385***	-0.0478
	(0.0033)	(0.0098)	(0.0291)	(0.0437)
STEM broader definition	-0.0141***	0.1355***	-0.1318***	-0.0648
	(0.0034)	(0.0102)	(0.0291)	(0.0432)
STEM stricter definition	-0.0091	0.0555*	-0.1604***	-0.0536
	(0.0067)	(0.0310)	(0.0244)	(0.1030)
Full-time workers	-0.0150***	0.1146***	-0.0779***	-0.0135
	(0.0039)	(0.0101)	(0.0274)	(0.0372)
More than 10 years in US	-0.0179***	0.1047***	-0.1395***	-0.0040
	(0.0044)	(0.0131)	(0.0371)	(0.0454)
Natives	-0.0076***	0.0828***	-0.1669***	-0.0092
	(0.0014)	(0.0052)	(0.0195)	(0.0290)
Naturalized citizens	-0.0089*	0.1157***	-0.0947**	-0.0179
	(0.0051)	(0.0160)	(0.0414)	(0.0440)
Non-citizens	-0.0230***	0.1567***	-0.1753**	-0.1475*
	(0.0060)	(0.0175)	(0.0696)	(0.0846)
Industry Fixed Effects	Yes	Yes	Yes	Yes
MSA Fixed Effects	Yes	Yes	Yes	Yes

Notes: The omitted base group includes college graduates with bachelor's degrees in all non-STEM fields. All regressions include individual controls provided upon request. Standard errors in parentheses are robust to heteroskedasticity and clustered by MSA.

The fifth row of Table 2.7 restricts the sample to foreigners who stay in the U.S. more than 10 years. Considering a quite long time spent in the host country, immigrants could be very different from the new comers from every aspect, including language skills, institutional constraints, relying on current income, etc. However, there is no qualitative change in the estimated coefficients. In order to understand this more clearly, natives are taken as reference. Those who stay in the U.S. more than 10 years could behave more like natives rather than the new comers. Nevertheless, the sixth row indicates that this is not the case since the estimates for natives is quite different from immigrants, no matter how long they stay in the U.S. This is somewhat opposite to our common sense.

One possible explanation is the citizenship status. Although some immigrants who have stayed in this country for considerably long time, they may not have citizenship. This institutional constraint could be a major reason that they still behave like new immigrants. The estimates for naturalized citizens and non-citizens support this idea. According to the estimates, naturalized citizens and natives are very similar, but non-citizen STEM graduates are much less likely to be self-employed. For the income regression, notice that only the interaction term for non-citizens is significantly negative at the 10% level, which could partly explain the much lower self-employed coefficient non-citizens have. Therefore, according to this exercise, income differentials and self-employment differentials of foreign college graduates could be jointly determined by their citizenship status.

5 Conclusion

This paper uses 1% IPUMS ACS 2013 data to study the self-employment differential between foreign STEM graduates and non-STEM graduates. The empirical results show that the differential is still quite substantial after controlling for demographic and socioeconomic characteristics, educational attainments, country groups of origin, language ability, industry fixed effects, and MSA fixed effects. Furthermore, self-employment differentials across broad major groups and detailed majors are examined. We find that the majors in Health, Social Science, and Liberal Arts are highlighted with relative high self-employment rates, while most majors in STEM fields are with lowest rates.

We try to explain the self-employment differentials through the differences in incomes between self-employed and salaried foreign college graduates. If foreign college graduates care more about current incomes, the higher incomes of salaried jobs, the lower probability of being self-employed. The lower incomes of self-employment, the lower self-employment rates. Our empirical results confirm these relationships. Foreign STEM graduates are less likely to be self-employed, since they could earn significantly more in salaried jobs, but the income advantage disappears when they shift to self-employment.

The sensitivity analysis provides very interesting results. We find that the old immigrants who stay in the U.S. for over 10 years still behave similarly as new comers on self-employment decision. This is partly because citizenship matters. The results show naturalized citizens and natives are very similar on self-employment decision, but non-citizens have much lower possibilities to be self-employed. Therefore, income differentials is one of the factor explaining the low self-employment rates among foreign STEM graduates, and the institutional constraint is another.

Considering the descriptive nature of this paper, we could not draw very strong conclusions. However, it is important for policymakers to know how immigrants with college degrees contribute to small business sector. Foreign STEM graduates are productive and innovative, but with very low self-employment rates. Thus, on one hand, policy makers could consider lower the immigration barrier for graduates in non-STEM fields with high self-employment rates, or at least reduce the institutional discrimination between STEM and non-STEM graduates. On the other hand, target based subsidies and/or tax benefits could be offered to the startups co-funded by foreign STEM graduates and those who are educated in fields with high self-employment rates.

CHAPTER III

HUMAN CAPITAL EXTERNALITIES, AGGLOMERATION, AND HOURS WORKED

1 Introduction

Firms locate in clusters because they can have net benefits for doing so. Learning from others is one of the advantages. However, it is rare in the literature to test the relationship between human capital externality and agglomeration directly. This paper uses unique instrumental variables to estimate the causality from both directions, namely, from human capital spillover to agglomeration, and also from agglomeration economies to human capital stock. After the causality is established, it can be used to analyze the relationship between human capital externality and hours worked given that the relationship between agglomeration and hours worked is known from the literature.

This paper contributes to the literature in several aspects. First, the empirical results show that human capital externality has much more causal impact on agglomeration than the reverse. The relationship between human capital externality and urbanization is much stronger than that with localization for both directions of causality.

Second, human capital externality does not affect hours worked of the self-employed, since it is localization not urbanization that has causal impact on hours worked, and the relationship between human capital spillover and localization is trivial for the self-employed. Human capital externality has impact on hours worked of employees, because the relationship between human capital and localization is significant. Based on these results, local governments could consider two types of policies to increase economic activity. One is how to attract more workers with higher educational backgrounds. The other is how to lower the probabilities for such workers and local university graduates moving away. Then combining with the policies suggested in Chapter I, regional economies could get more benefits from clusters.

Third, this paper uses three valid instrumental variables. Minimum distance from work PUMA centroid to land-grant university is used to instrument for human capital stock, which is a significant improvement of the early version of the instrument in the literature at a larger scope of geographic level. Following Chapter 1, minimum distance from work PUMA centroid to US shoreline and estimated industry share in 1930 are used to instrument for agglomeration measures.

Lastly, the wage rate is used instead of hours worked in the robustness check. The results show that the mechanism between human capital externality and agglomeration also works for the relationship between human capital externality and wages.

2 Literature Review

2.1 Agglomeration affects hours worked

The relationship between agglomeration and hours worked is overlooked in the literature. The few existing studies show that only localization is positively correlated with hours worked. Urbanization has nothing to do with work intensity.

Rosenthal and Strange (2008a) document the positive relationship between localization and hours worked of professional workers. They argue that this positive relationship mainly goes through the urban rat race. Urban rat race is a mechanism that under competitive circumstance, workers tend to work more hours in order to signal their supervisors that they are hardworking, which could increase the probability of getting ahead.

Chapter 1 uses instrumental variable estimation to build the causal relationship between agglomeration and hours worked for the self-employed. We find that only localization increases work intensity of the self-employed, while the agglomeration wage effect only comes from urbanization. The mechanism is different from the urban rat race for employees. Localization brings more competition within industries, which causes longer hours worked. Concurrently, specialization also decreases competition across industries, and thus urbanization has no causal impact on hours worked.

2.2 Human capital externality and Agglomeration

Human capital externality is an important source of agglomeration, which is a consensus in the literature. However, few studies really test the relationship between human capital externality and agglomeration.

Most of the studies focus on indirect association between human capital externality and agglomeration, including wages, crimes, and politics. Moretti (2004a) reviews the literature on the social return of human capital, and conclude that human capital externality increases productivity, reduces criminal participation, and improves voters' political behavior. Rosenthal and Strange (2004) is also a good survey of the empirical literature on knowledge spillover as a micro-foundation of agglomeration. Duranton and Puga (2004) study the theoretical micro-foundations of agglomeration. Learning is an important mechanism.

Nonetheless, how human capital externality is empirically correlated with agglomeration still remains unknown. This paper uses instrumental variables for both human capital and agglomeration to examine their relationship from both directions. That is, not only assuming that human capital externality causes the formation of agglomeration, but also assuming agglomeration will affect human capital stock.

Besides, the relationship between human capital externality and work intensity is overlooked as well. Most studies focus on wage regressions. Moretti (2004b) uses longitudinal and repeated cross-sectional data to estimate the social return to higher education. The results show that human capital externality raises wages, and less educated workers receive more external benefits.

Fu (2007) studies the micro-foundations of human capital externality. He finds that human capital externality could go through four channels: depth of human capital stock, Marshallian labor market externalities, Jacobs labor market externalities, and thickness of the local labor market. Actually, the first channel is measured by the share of working population who have at least a college degree, which is the mostly used measure of human capital. The second channel is related to specialization and peer competition, and thus could be measured by localization as well. The third and fourth channels are very related to urbanization. His results show that these four types of mechanism are significant at the census block level, and attenuate at different speeds with distance.

Rosenthal and Strange (2008b) estimate the relationship between agglomeration, human capital externality and wages. They find that agglomeration is positively associated with wage; the benefits of agglomeration are driven by human capital spillovers; and the benefits attenuate with distance.

The only study related to the relationship between human capital externality and hours worked is Winters (2013). He finds that human capital level externality has a positive effect on the probability of labor force participation and employment in US metropolitan areas.

According to the literature above, agglomeration affects hours worked, and human capital externality is a source of agglomeration. It is interesting to look at the relationship between human capital externality and hours worked. It is expected that human capital externality may affect hours worked as well, since the effects of human capital externality and agglomeration share a lot of common features, including spillover effects on wages, productivity, attenuation with distance, etc. However, it is perhaps not the case as well. Since only localization could affect hours worked but not urbanization, it is necessary to examine the detailed relationship between human capital spillover and different types of agglomeration economies.

Therefore, this paper will examine the causality from human capital spillover to hours worked, and more importantly, see why it is the case. The next section will discuss our empirical framework. Section 4 introduces the data and variables. Section 5 presents the empirical results. The last section concludes.

3 Empirical Framework

3.1 Human Capital Externality and Hours Worked

The empirical analysis starts with looking at the relationship between human capital externality and hours worked. The baseline model is:

$$\begin{aligned} \log\left(Hours\ Worked_{ipjc}\right) \\ &= \alpha \log\left(College\ Share_p\right) + \boldsymbol{X}_{ipjc}\boldsymbol{\beta} + \boldsymbol{Amenity}_p\boldsymbol{\gamma} \\ &+ \mu_j + \varphi_c + \varepsilon_{ipjc} \end{aligned} \tag{9}$$

where *i*, *p*, *j*, *c* indexes individual observations, work PUMA, industries, and work MSAs, respectively.[35] $College\ Share_p$ is the share of working aged population who own bachelor's degree and above at work PUMA level. $\boldsymbol{X}_{ipjc}$ is a set of demographic controls. μ_j is industry fixed effect. φ_c is work MSA fixed effect. ε_{ipjc} is an error term.

According to the related literature, agglomeration is associated with human capital externalities (Moretti, 2004a; Duranton and Puga, 2004; Rosenthal and Strange, 2004), and also correlated with hours worked (Rosenthal and Strange, 2008; Chapter 1). Thus, agglomeration measures are included in the regression as a robustness check. The model is:

[35] All the empirical steps use personal sampling weights (perwt in IPUMS) to ensure the results are nationally representative.

$$\log\left(Hours\ Worked_{ipjc}\right) = \alpha \log\left(College\ Share_p\right) + \beta \log\left(Urbanization_p\right) + \gamma \log\left(\frac{Localization_{pj}}{Urbanization_p}\right) + \boldsymbol{X}_{ipjc}\boldsymbol{\theta} + \boldsymbol{Amenity}_p\boldsymbol{\delta} + \mu_j + \varphi_c + \varepsilon_{ipjc} \tag{10}$$

where $Urbanization_p$ is the population density of a work PUMA. $Localization_{pj}$ is measured by the industry-specific employment density of a work PUMA. Specifically, the employment in industry *j* is calculated for work PUMA *p*. Then $Localization_{pj}$ is the quotient of the industry-specific employment and the geographic area of the work PUMA *p*. Following Chapter 1, the localization measure is modified by urbanization to mitigate the collinearity issue.[36]

In order to take care of the potential endogeneity on human capital, an instrumental variable is employed. In the literature, one popular instrument is the presence of a land-grant college (Moretti, 2004; Iranzo and Peri, 2009; Winters, 2013). However, this instrument has little variation at the work PUMA level, which makes it a weak instrument. One feasible compromise is restricting the sample size with the cost of losing available information. This paper comes up with a better alternative modifying the early version of the instrument. The instrument used in this paper is the minimum distance from a work PUMA centroid to a land-grant college, which is less likely to suffer from the no variation problem. As documented in the literature, the land-grant college is a good instrument for college share because of its fair correlation with human

[36] As in Chapter I, I will use “localization” as referring to “relative localization” in the remainder of this paper to indicate the effect of the term $\log\left(\frac{Localization_{pj}}{Urbanization_p}\right)$.

capital measure, and relative exogeneity since the determination of the locations was a century ago and pretty random.[37]

3.2 Human Capital Externality and Agglomeration

Human capital externality is an important source of agglomeration economies. Agglomeration affects hours worked (Rosenthal and Strange, 2008; Chapter 1). Thus, it is natural to take a look at the role of agglomeration in the relationship between human capital externality and hours worked. To our best knowledge, only the one-way relationship between human capital spillover and agglomeration is usually studied in the literature. In this paper, we also study the reverse causality from agglomeration to human capital measure in order to guarantee the completeness of our inference. The baseline regression models are estimated by OLS as following:

$$\log\left(Urbanization_p\right) = \alpha \log\left(College\ Share_p\right) + \boldsymbol{X}_p\boldsymbol{\theta} + \boldsymbol{Amenity}_p\boldsymbol{\delta} + \varphi_c + \varepsilon_{pc} \tag{11}$$

$$\log\left(\frac{Localization_{pj}}{Urbanization_p}\right) = \alpha \log\left(College\ Share_p\right) + \boldsymbol{X}_p\boldsymbol{\theta} + \boldsymbol{Amenity}_p\boldsymbol{\delta} + \varphi_c + \varepsilon_{pc} \tag{12}$$

$$\log\left(College\ Share_p\right) = \alpha \log\left(Urbanization_p\right) + \beta \log\left(\frac{Localization_{pj}}{Urbanization_p}\right) + \boldsymbol{X}_p\boldsymbol{\theta} + \boldsymbol{Amenity}_p\boldsymbol{\delta} + \varphi_c + \varepsilon_{pc} \tag{13}$$

[37] This instrumental variable may also have some limitations. See a discussion in Winters (2013).

where $\boldsymbol{X}_p$ is a set of covariates of a work PUMA. Model (3) and (4) test the relationship from human capital externality to agglomeration. Model (5) examines the same relationship from the reverse direction.

However, simultaneity is very likely to bias the estimates. Thus, minimum distance from work PUMA centroid to land-grant college is used to instrument for human capital externality in Models (3) and (4). Following Chapter 1, minimum distance from work PUMA centroid to US shoreline and estimated industry share in 1930 are employed to instrument for agglomeration measures in Model (5). After examining the causality between human capital externality and agglomeration from both directions, we are able to analyze the mechanism of the relationship between human capital externality and hours worked.

3.3 Robustness Check

To test whether the mechanism is robust to other outcomes, regression-adjusted incomes are constructed for each work PUMA by the regression:

$$\ln(Hourly\ Income)_{ipj} = \boldsymbol{X}_{ipj}\boldsymbol{\beta} + \mu_j + \omega_p + \varepsilon_{icd} \tag{14}$$

where ω_p is work PUMA fixed effects, which is used as the regression-adjusted average log hourly income of a work PUMA. This regression is estimated for the self-employed and the employed separately to obtain regression-adjusted hourly income $\omega_c^{Self-Employed}$ for the self-employed, and $\omega_c^{Employed}$ for the employed. Then $\omega_c^{Self-Employed}$ will be used in the employed sample, $\omega_c^{Employed}$ will be used in the self-employed sample to ensure the exogeneity and capture the spillover aspect of human capital externality. Finally, the causality between human capital externality and incomes will be used to verify the prediction of the mechanism.

4 Data and Variables

4.1 Sample and Sub-samples

The data is retrieved from the 5% IPUMS 2000 sample covering the lower 48 states. The sample is restricted to full-time workers[38] aged from 30 to 59[39], and divided into two educational groups: high school degree or below, and college degree or above. Each educational group is subdivided into three age groups: 30 to 39, 40 to 49, and 50 to 59.

Educational attainment, a dummy of the presence of children, dummies of marital status, polynomial of age, dummies of race, years of residency in the US, and commute time are included as the demographic controls.

4.2 Human Capital Measure

Fu (2007) shows that human capital spillovers go through four channels. One of the important channels is the depth of human capital stock in the local labor market, which is measured by the share of working aged population who have at least bachelor's degree.[40] This paper uses this measure as the proxy of human capital externality.

---

[38] Working time is 35 hours or more per week.

[39] People aged 30 - 59 cover about 70% of the full-time workers.

[40] This paper uses college degree share or college share short for the share of working aged population who have at least bachelor's degree.

4.3 Land-Grant Universities

In order to construct the instrument, the land-grant universities should be defined. This paper combines the list of land-grant universities in the Appendix of Nevins (1962), and also the list of 1862 and 1890 land-grant colleges and universities from National Institute of Food and Agriculture (NIFA) of United States Department of Agriculture (USDA). Based on the difference between these two lists, this paper identifies 67 land-grant universities.[41] The coordinates of the official address of the land-grant universities are imported into GIS software as the location points of the universities. Then the minimum distance from work PUMA centroid to a land-grant university is calculated by the GIS software.

4.4 Agglomeration Measures

As discussed in Section 2, Marshallian labor market externality, Jacobs labor market externality, and the thickness of the local labor market summarized by Fu (2007) are very similar with the agglomeration measures. The relative localization measure in Chapter 1 could capture the Marshallian labor market externality. The urbanization measure in Chapter 1 is very close to the concepts of Jacobs labor market externality and the density of the local labor market. Therefore, this paper uses the urbanization and relative localization measures of Chapter 1.[42]

4.5 Dependent Variables

[41] The full list is provided in the Appendix.
[42] See Chapter 1 for the details.

Usual hours worked per week in the previous year (uhrswork in IPUMS) is used as the measure of hours worked. Earned income (INCEARN in IPUMS) is used as the annual income measure. Hourly income is the quotient of earned income and hours worked in the previous year.

4.6 Amenities

Following Chapter 1, amenities are extracted from different sources and constructed at the work PUMA level, including violent crime, property crime, precipitation, January temperature, July temperature, elevation, minimum distance to the nearest river or lake, heating degree days, cooling degree days, dew points, direct solar irradiance, and four dummies for coastal work PUMAs of Atlantic Ocean, Pacific Ocean, Gulf of Mexico, and Great Lakes.[43] Table 3.1 shows the summary statistics.

Table 3.1: Summary Statistics

Variable	No. Obs	Mean	Std. Dev.	Min	Max
Hours Worked	1,850,145	45.022	8.857	35.000	99.000
Log (Hours Worked)	1,850,145	3.791	0.172	3.555	4.595
Hourly Income	1,850,145	21.371	34.784	-285.714	8150.000
Log (Hourly Income)	1,844,642	2.746	0.754	-7.160	9.006
College Share	1,850,145	0.273	0.095	0.102	0.595
Log (College Share)	1,850,145	-1.357	0.340	-2.283	-0.519
Localization	1,850,145	137.863	705.882	0.000	8455.254
Log (Localization)	1,850,145	1.799	2.397	-8.987	9.043
Urbanization	1,850,145	3731.385	11513.840	1.637	66942.260
Log (Urbanization)	1,850,145	6.211	1.975	0.493	11.112
Minimum Distance to Land-Grant	1,850,145	97.683	88.850	0.619	383.869
Log (Minimum Distance to Land-Grant)	1,850,145	4.135	1.085	-0.480	5.950
Minimum Distance to Coastline	1,850,145	139.165	174.596	0.023	817.575
Log (Minimum Distance to Coastline)	1,850,145	3.822	1.850	-3.772	6.706
Imputed Industry Share in 1930	1,850,145	0.032	0.087	0.000	1.000
Log (Industry Share in 1930)	1,850,145	-4.561	1.476	-12.368	0.000
High School and Less	1,850,145	0.545	0.498	0.000	1.000

[43] See Chapter 1 for the details.

College and More	1,850,145	0.455	0.498	0.000	1.000
Age	1,850,145	43.146	8.019	30.000	59.000
Log (Commute Time)	1,792,663	2.960	0.848	0.000	5.159
Children Present	1,850,145	0.546	0.498	0.000	1.000
Female	1,850,145	0.409	0.492	0.000	1.000
Marital Status					
Married	1,850,145	0.671	0.470	0.000	1.000
Married, Spouse Absent	1,850,145	0.018	0.133	0.000	1.000
Separated	1,850,145	0.025	0.157	0.000	1.000
Divorced	1,850,145	0.135	0.341	0.000	1.000
Widowed	1,850,145	0.014	0.116	0.000	1.000
Never Married	1,850,145	0.138	0.345	0.000	1.000
Race					
White	1,850,145	0.799	0.401	0.000	1.000
African American	1,850,145	0.092	0.289	0.000	1.000
American Indian or Alaska Native	1,850,145	0.006	0.080	0.000	1.000
Chinese	1,850,145	0.013	0.111	0.000	1.000
Japanese	1,850,145	0.003	0.051	0.000	1.000
Other Asian or Pacific Islander	1,850,145	0.027	0.163	0.000	1.000
Other Race	1,850,145	0.042	0.202	0.000	1.000
Two Major Races	1,850,145	0.016	0.126	0.000	1.000
Three or More Major races	1,850,145	0.001	0.028	0.000	1.000
Amenity					
Log (Violent Crime)	1,846,529	6.790	1.685	1.946	10.380
Log (Property Crime)	1,847,489	7.947	1.504	2.639	12.000
Log (Precipitation)	1,850,145	8.953	0.458	6.677	9.998
Log (Dew Points)	1,850,145	7.168	0.571	-10.735	7.855
Log (January Temperature)	1,850,145	2.749	0.580	-13.356	3.540
Log (July Temperature)	1,850,145	3.182	0.137	2.608	3.499
Log (Heating Degree Days)	1,850,145	7.557	0.821	3.664	8.616
Log (Cooling Degree Days)	1,850,145	7.665	0.411	6.157	8.556
Log (Elevation)	1,850,145	8.403	0.792	-7.953	8.904
Log (Solar Irradiance)	1,850,145	1.505	0.204	1.129	2.069
Log (Minimum Distance to River and Lake)	1,850,145	2.117	1.205	-2.052	4.432
Atlantic Work PUMA	1,850,145	0.171	0.376	0.000	1.000
Great Lake Work PUMA	1,850,145	0.057	0.232	0.000	1.000
Gulf Work PUMA	1,850,145	0.048	0.213	0.000	1.000
Pacific Work PUMA	1,850,145	0.117	0.322	0.000	1.000

Notes: All summary statistics are adjusted by personal weight to ensure the national representative. Education, years of residency in the U.S., industry, work PUMA, and work MSA are not included for space conservation.

5 Empirical Results

This section starts with looking at the relationship between human capital externality and hours worked for the employed and the self-employed. Instrumental variable estimation then will be used to establish the causality between the two variables. This paper will focus on the mechanism between human capital spillovers and agglomeration economies. Thus, causality will be analyzed from both directions.

5.1 Human Capital Externality and Hours Worked

In order to have a sense of potential bias, OLS is employed to test the correlation between human capital externality and hours worked first. Panel A of Table 3.2 reports the OLS results of model (1) for the employed. Almost all the estimates are statistically significant, but it indicates different patterns for the lower educated group and the higher educated group. The higher college degree share is associated with lower hours worked for the lower educated group, while higher college degree share is correlated with higher hours worked for higher educated employees.

For the reasons discussed in Section 3, agglomeration measures are added to the model. Panel B of Table 3.2 shows the results. However, the inclusion of urbanization and relative localization does not change the estimates on college degree share much. Therefore, besides, concerning the endogeneity of agglomeration measures, the preferred specification only controls for the human capital measure.

Now, instrumental variable estimation is used to take care of the potential endogeneity issue. Panel C of Table 3.2 reports the 2SLS results. Most estimates are insignificantly different from zero. The lower part of Panel C reports the first stage estimation results. All the first stage estimates are negative and significant, indicating that

Table 3.2: Hours Worked and Human Capital Externality for Employees

	Dependent Variable: Log (Hours Worked)					
	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
A. Ordinary Least Squares						
Log (College Share)	-0.0020	-0.0053**	-0.0098***	0.0246***	0.0171***	0.0193***
	(0.0024)	(0.0021)	(0.0023)	(0.0052)	(0.0038)	(0.0044)
B. Ordinary Least Squares						
Log (College Share)	-0.0042	-0.0040	-0.0092***	0.0243***	0.0230***	0.0286***
	(0.0029)	(0.0025)	(0.0028)	(0.0053)	(0.0040)	(0.0048)
Log (Urbanization)	0.0005	-0.0009*	-0.0004	-0.0011	-0.0023***	-0.0027***
	(0.0005)	(0.0005)	(0.0006)	(0.0008)	(0.0007)	(0.0007)
Log(Localization/Urbanization)	0.0018***	0.0012***	0.0004	0.0020**	-0.0007	-0.0027***
	(0.0004)	(0.0003)	(0.0004)	(0.0008)	(0.0007)	(0.0006)
C. Two Stage Least Squares						
Log (College Share)	0.0148	-0.0191**	-0.0110	0.0169	0.0060	0.0278*
	(0.0108)	(0.0095)	(0.0105)	(0.0133)	(0.0115)	(0.0144)
First Stage						
Log (Distance to Land Grant University)	-0.0701***	-0.0719***	-0.0711***	-0.0768***	-0.0769***	-0.0790***
	(0.0122)	(0.0119)	(0.0121)	(0.0159)	(0.0143)	(0.0137)
First Stage F Statistics	32.880	36.352	34.678	23.399	28.741	33.475
Endogeneity	2.531	2.325	0.014	0.303	0.931	0.342
	[0.1116]	[0.1273]	[0.9057]	[0.5823]	[0.3347]	[0.5587]
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * p < 0.1, **p<0.05, *** p < 0.01.

the further from the land grant universities, the lower the college degree shares. All the first stage F-statistics are more than 20, which guarantees that the instrumental variable is not a weak instrument. However, the endogeneity tests fail to reject that college share is exogenous. Considering the lesser efficiency of 2SLS, OLS estimates are more reliable,

Table 3.3: Hours Worked and Human Capital Externality for the Self-Employed

Dependent Variable: Log (Hours Worked)	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
A. Ordinary Least Squares						
Log (College Share)	-0.0154*	0.0001	-0.0132	0.0037	0.0107	0.0020
	(0.0083)	(0.0068)	(0.0085)	(0.0105)	(0.0078)	(0.0078)
B. Ordinary Least Squares						
Log (College Share)	-0.0191*	0.0075	-0.0109	-0.0053	0.0058	0.0002
	(0.0103)	(0.0079)	(0.0098)	(0.0116)	(0.0096)	(0.0099)
Log (Urbanization)	-0.0019	-0.0066***	-0.0042**	-0.0010	-0.0015	-0.0037*
	(0.0020)	(0.0016)	(0.0019)	(0.0024)	(0.0020)	(0.0022)
Log(Localization/Urbanization)	0.0131***	0.0112***	0.0107***	0.0073***	0.0060***	0.0081***
	(0.0017)	(0.0015)	(0.0015)	(0.0015)	(0.0013)	(0.0013)
C. Two Stage Least Squares						
Log (College Share)	-0.0231	-0.0132	-0.0224	0.0371	0.0079	-0.0313
	(0.0396)	(0.0315)	(0.0436)	(0.0470)	(0.0377)	(0.0391)
First Stage						
Log (Distance to Land Grant University)	-0.0734***	-0.0770***	-0.0678***	-0.0809***	-0.0783***	-0.0769***
	(0.0123)	(0.0121)	(0.0116)	(0.0149)	(0.0134)	(0.0137)
First Stage F Statistics	35.318	40.384	34.196	29.330	33.882	31.302
Endogeneity	0.040	0.195	0.046	0.510	0.006	0.765
	[0.8414]	[0.6589]	[0.8296]	[0.4753]	[0.9402]	[0.3816]
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * $p < 0.1$, ** $p<0.05$, *** $p < 0.01$.

indicating that higher human capital externalities make lower educated employees to work less, while higher educated employees to work more.

Table 3.3 replicates Table 3.2 for the self-employed. The difference between these two samples is that there is little correlation between human capital externality and hours worked of the self-employed by OLS and instrumental variable estimation. The endogeneity tests indicate that the OLS results are more reliable.

To sum up, according to Table 3.2 and Table 3.3, human capital externalities affect hours worked of employees but not the self-employed. In the following sections, we want to look at the possible mechanism behind the results.

5.2 Human Capital Externality and Agglomeration

According to the literature, human capital externality is an important source of agglomeration economies. Under this relationship, however, if agglomeration has causal impact on hours worked, why does not human capital externality affect hours worked of the self-employed? Why does the impact exist for employees? In order to answer these questions, it is necessary to examine the causality between human capital externality and agglomeration from both directions.

First, following the literature, we test human capital spillover as a source of agglomeration. To have a sense of endogeneity, OLS and instrumental variable estimation will be used in succession.

Table 3.4 reports the estimated correlation between human capital externality and urbanization by OLS for different samples. The first column shows the elasticity between human capital externality and urbanization is 1.89, which is pretty substantial. Columns (2) and (3) report the estimates by types of employment, indicating that the estimated elasticity for the self-employed is a little bit larger than that for employees, but the difference is not statistically substantial.

Table 3.5 replicates Table 3.4 but for relative localization. The results are similar with those in Table 3.4, just less in magnitudes and significance. The elasticity between human capital externality and relative localization is about 1 for the full sample.

Table 3.4: Human Capital Externality and Urbanization

Dependent Variable: Log (Urbanization)			
	(1)	(2)	(3)
	Full Sample	Self-Employed	Employed
Log (College Share)	1.8950***	1.9418***	1.8878***
	(0.2320)	(0.2167)	(0.2336)
Amenities	Yes	Yes	Yes
Work MSA Fixed Effects	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work state. * $p < 0.1$, **$p<0.05$, *** $p < 0.01$.

Table 3.5: Human Capital Externality and Localization

Dependent Variable: Log (Localization/Urbanization)			
	(1)	(2)	(3)
	Full Sample	Self-Employed	Employed
Log (College Share)	1.0012**	1.0135**	0.9983**
	(0.3812)	(0.4031)	(0.3797)
Amenities	Yes	Yes	Yes
Work MSA Fixed Effects	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work state. * $p < 0.1$, **$p<0.05$, *** $p < 0.01$.

Table 3.6 provides the estimated elasticity reversely between agglomeration and human capital externality with respect to Tables 3.4 and 3.5. Compared with Tables 3.4 and 3.5, the estimates are less in magnitudes. The estimates on relative localization are even smaller than those on urbanization and less significant. For the full sample, the elasticity between urbanization and college degree share is 0.12, and 0.01 for relative localization and college degree share.

Therefore, according to the results from Table 3.4 to Table 3.6, the association between human capital externality and agglomeration is more likely to start from human capital externality, and the association is stronger for urbanization than localization. However, a strong conclusion cannot be made by the OLS estimates because of potential endogeneity. Thus, instrumental variable estimation is used to analyze causality.

Table 3.7 reports the instrument variable estimation results for the full sample, the self-employed, and employees, assuming the causal relationship is running from human capital externality to agglomeration. For the full sample, although the second stage estimates are positive and significant, the relationship between human capital spillover and urbanization is substantially stronger than that for localization. Since the first stages are identical for the urbanization and relative localization regression, the first stage estimates are same. As in Table 3.2 and Table 3.3, the estimated coefficient on the instrument is negative and highly significant. The first stage F-statistic is 27.794, which is large enough to guarantee that the instrument is not weak.

In order to investigate why the relationship between human capital externalities and hours worked is different for the self-employed and employees, Table 3.7 also provides the estimates for these two samples. The estimates for urbanization are similar for the two samples, which means the human capital externalities have similar causal effects on urbanization for the self-employed and employees. However, human capital externalities have little impact on localization for the self-employed compared with employees, and this might be the reason why human capital externalities affect hours worked of employees rather than the self-employed. Given localization has a positive impact on hours worked, nevertheless, since human capital externalities do not contribute to the formation of localization economies for the self-employed, the causality could not go through this channel. On the other hand, human capital spillovers affect localization for employees, and localization has impact on hours worked, thus human capital externalities affect hours worked for employees.

Table 3.8 gives the instrumental variable estimation results of the reverse relationship, assuming the causality is from agglomeration to human capital. For the full sample, the elasticity between urbanization and college degree share is 0.1940. The estimate on relative localization is insignificant different from zero, indicating there is no causal impact of localization on human capital stocks. Therefore, in this direction, localization does not affect human capital stocks, so that it cannot affect hours worked through the variation in human capital externalities. The estimates for the self-employed and employees are very similar as the full sample. Relying on the results of Tables 3.7 and 3.8, it implies that the causality is mostly running from human capital to agglomeration, and confirms the statement in the literature that human capital externality is a source of agglomeration economy. Yet, urbanization still has some causal impact on human capital, although it is much smaller than the reverse relationship.

Table 3.6: Agglomeration and Human Capital Externality

Dependent Variable: Log (College Share)			
	(1)	(2)	(3)
	Full Sample	Self-Employed	Employed
Log(Urbanization)	0.1216***	0.1202***	0.1218***
	(0.0157)	(0.0133)	(0.0160)
Log(Localization/Urbanization)	0.0135**	0.0170**	0.0134**
	(0.0066)	(0.0083)	(0.0066)
Amenities	Yes	Yes	Yes
Work MSA Fixed Effects	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work state. * p < 0.1, **p<0.05, *** p < 0.01.

Table 3.7: Human Capital Externality and Agglomeration

	Log(Urbanization)			Log(Localization/Urbanization)		
	Full Sample	Self-Employed	Employed	Full Sample	Self-Employed	Employed
Log (College Share)	3.2834***	3.7311***	3.2263***	0.5947**	0.2608	0.6233**
	(0.7857)	(0.7903)	(0.7850)	(0.2982)	(0.3090)	(0.3065)
First Stage						
Log (Distance to Land Grant University)	-0.0804***	-0.0784***	-0.0806***	-0.0804***	-0.0784***	-0.0806***
	(0.0152)	(0.0139)	(0.0154)	(0.0152)	(0.0139)	(0.0154)
First Stage F Statistics	27.794	31.749	27.448	27.794	31.749	27.448
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work state. * p < 0.1, **p<0.05, *** p < 0.01.

Table 3.8: Agglomeration and Human Capital Externality

	Log(College Share)		
	Full Sample	Self-Employed	Employed
Log(Urbanization)	0.1940***	0.2006***	0.1931***
	(0.0739)	(0.0710)	(0.0743)
Log(Localization/Urbanization)	-0.0031	0.0012	-0.0034
	(0.0043)	(0.0069)	(0.0042)
First Stage			
Urbanization			
Log (Distance to Shoreline)	-0.4083***	-0.3698***	-0.4123***
	(0.0778)	(0.0798)	(0.0777)
Log (Industry share in 1930)	0.0277***	0.0462***	0.0266***
	(0.0060)	(0.0100)	(0.0058)
Localization/Urbanization			
Log (Distance to Shoreline)	-0.1518	-0.1195	0.1552
	(0.0983)	(0.0940)	(0.0986)
Log (Industry share in 1930)	0.4009***	0.4470***	0.3952***
	(0.0160)	(0.0275)	(0.0156)
Kleibergen-Paap rk Wald F statistic	16.758	12.464	17.257
Amenities	Yes	Yes	Yes
Work MSA Fixed Effects	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work state. * $p < 0.1$, ** $p<0.05$, *** $p < 0.01$.

5.3 Summary of the Mechanism

After the causality analysis between human capital and agglomeration from both directions, we could sum up the mechanism behind the causality between human capital externality and hours worked.

In order to have a clearer illustration, based on the results from Tables 3.2, 3.3, 3.7, and 3.8, Figures 3.1 and 3.2 are produced to summarize the mechanism for the self-employed and employees, respectively. This section will follow Figures 3.1 and 3.2, and go over the mechanism from two directions. The first direction is from human capital externality to agglomeration, the other direction is reverse.

Figure 3.1 shows that human capital externality strongly increases urbanization, but it has little impact on localization for the self-employed. According to the literature, urbanization has no causal impact on hours worked, localization significantly increases hours worked. Therefore, the causality from human capital externality to hours worked cannot run through the two channels. Thus, the causality of human capital externality on hours worked is insignificant for the self-employed. Figure 3.2, shows that human capital affects localization for employees as well. Therefore, the causality of human capital spillovers on hours worked of employees could go through the localization channel.

From the reverse direction, Figures 3.1 and 3.2 show the same thing that urbanization increases human capital, while localization does not. However, urbanization does not have a causal effect on hours worked, thus human capital should not have a causal impact on hours worked either. Meanwhile, localization has no causal impact on human capital. Therefore, the causality from localization to hours worked cannot pass through human capital externalities, and thus in this direction, human capital externalities have no causal impact on hours worked for both the self-employed and employees.

To sum up, the causality from human capital externality to hours worked of employees runs through the channel from human capital externality to localization. The relationship does not exist for the self-employed because the channels are blocked in both directions.

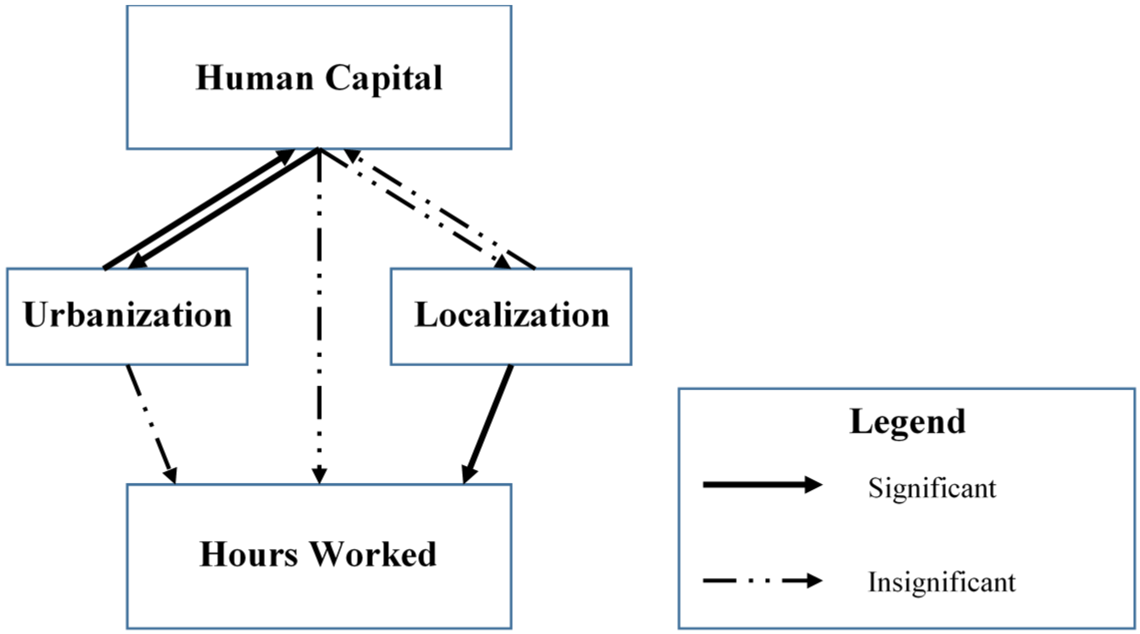

Figure 3.1: Mechanism behind the Relationship between Human Capital Externality and Hours Worked of the Self-Employed. Source: Author.

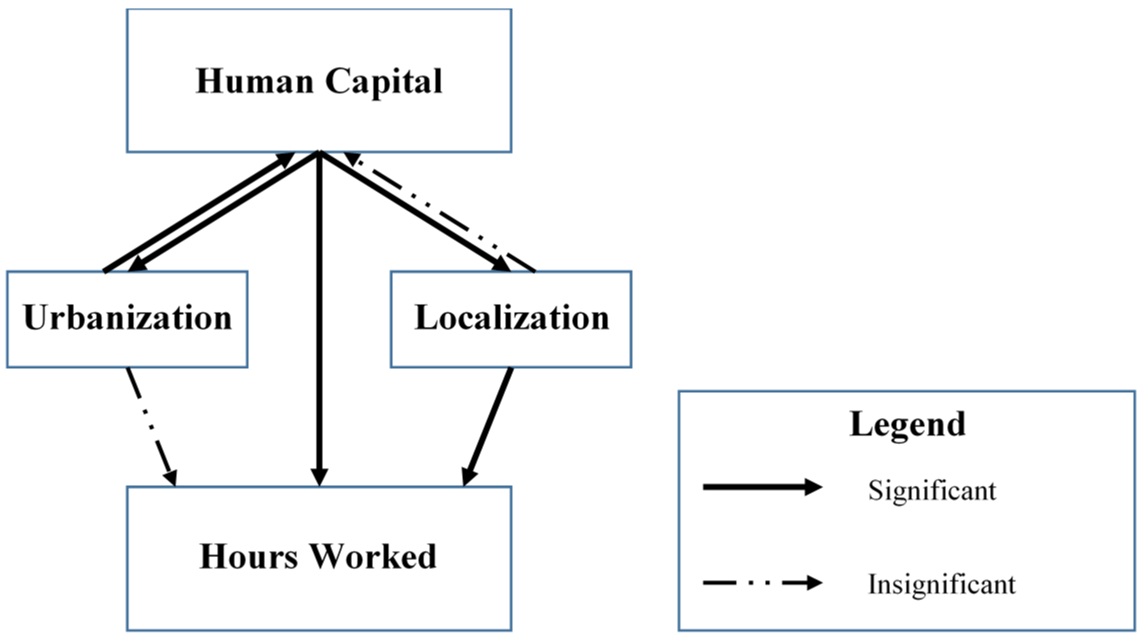

Figure 3.2: Mechanism behind the Relationship between Human Capital Externality and Hours Worked of Employees. Source: Author.

5.4 Robustness Check

A robustness check of the mechanism is examined to ensure the reliability of our analysis. Figures 3.3 and 3.4 present the robustness check. Hours worked in Figures 3.1

and 3.2 is changed to regression-adjusted wage at work PUMA level to see if the mechanism works as well.

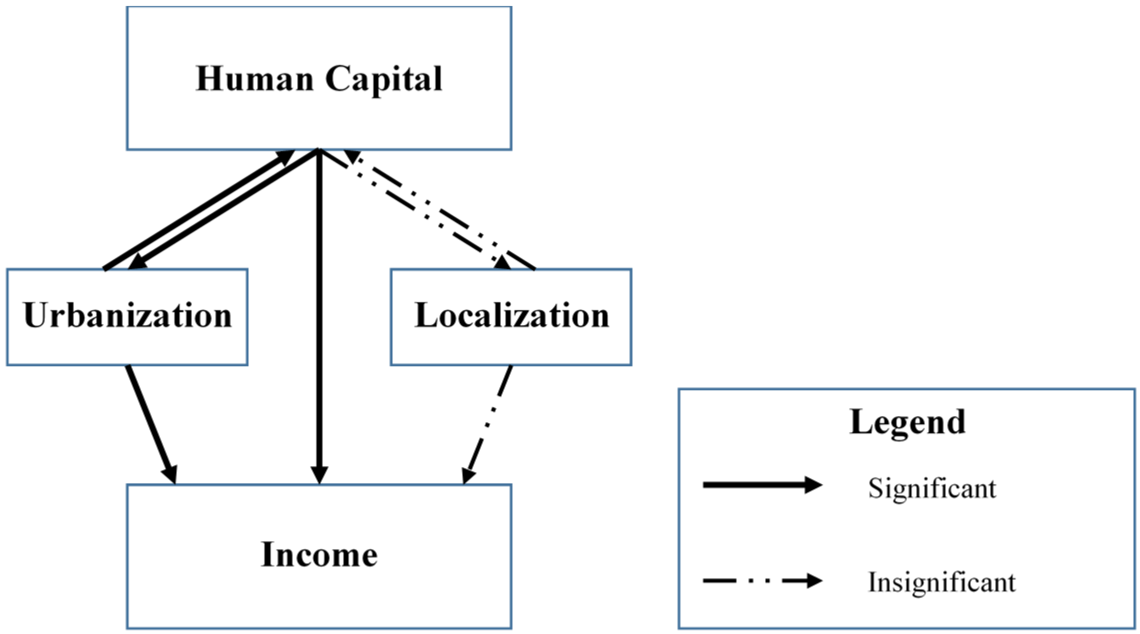

Figure 3.3: Mechanism behind the Relationship between Human Capital Externality and Incomes for the Self-Employed. Source: Author.

According to the literature, the agglomeration wage effect only comes from urbanization. Localization has no causal impact on wages. Based on the causal analysis between human capital and agglomeration from both directions, it can be easily inferred that human capital externality should also have causal impacts on wages since the channel between human capital and urbanization is unobstructed for both the self-employed and employees. In order to test whether the inference is valid, Tables 3.9 and 3.10 show the instrumental variable estimation results of the relationship between human capital externality and wages for the employed and the self-employed respectively. As discussed in Section 3, in order to have a relative exogenous wage measure, the income of the self-employed is used as the dependent variable in the regressions for employees; the wage of employees is used in the regression for the self-employed. All estimates are positive and significant, indicating that human capital externality has positive causal impact on wages as our prediction. Therefore, the mechanism analysis for the relationship between human capital externality and hours worked is valid.

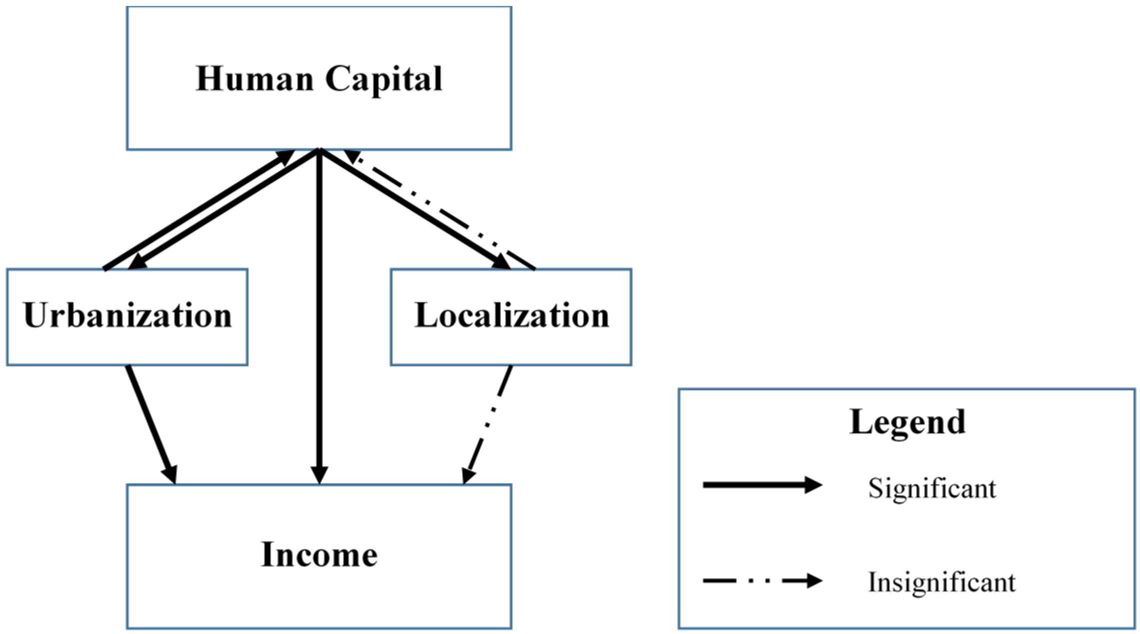

Figure 3.4: Mechanism behind the Relationship between Human Capital Externality and Incomes for Employees. Source: Author.

Table 3.9: Human Capital Externality and Wages for Employee Sample

Dependent Variable: Log (Adjusted Income for the Self-Employed)

	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
Log (College Share)	0.3284***	0.3173***	0.3137***	0.2770***	0.3009***	0.2934***
	(0.0824)	(0.0811)	(0.0827)	(0.1074)	(0.0959)	(0.0911)
First Stage						
Log (Distance to Land Grant University)	-0.0701***	-0.0719***	-0.0711***	-0.0768***	-0.0769***	-0.0790***
	(0.0122)	(0.0119)	(0.0121)	(0.0159)	(0.0143)	(0.0137)
First Stage F Statistics	32.880	36.352	34.678	23.399	28.741	33.475
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * p < 0.1, **p<0.05, *** p < 0.01.

Table 3.10: Human Capital Externality and Wages for the Self-Employed Sample

Dependent Variable: Log (Adjusted Income for Employee)

	High school and less			College and more		
	Age 30 - 39	Age 40 - 49	Age 50 - 59	Age 30 - 39	Age 40 - 49	Age 50 - 59
Log (College Share)	0.1927***	0.2025***	0.2051***	0.1695***	0.1787***	0.1784***
	(0.0334)	(0.0329)	(0.0381)	(0.0326)	(0.0340)	(0.0335)
First Stage						
Log (Distance to Land Grant University)	-0.0734***	-0.0770***	-0.0678***	-0.0809***	-0.0783***	-0.0769***
	(0.0123)	(0.0121)	(0.0116)	(0.0149)	(0.0134)	(0.0137)
First Stage F Statistics	35.318	40.384	34.196	29.330	33.882	31.302
Amenities	Yes	Yes	Yes	Yes	Yes	Yes
Industry and Work MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The other estimates are suppressed for space conservation. All regressions include individual controls listed in Table 3.1. Standard errors in parentheses are robust to heteroskedasticity and clustered by work PUMA. * p < 0.1, **p<0.05, *** p < 0.01.

6 Conclusion

This paper examines the causal relationship between human capital externality and hours worked. The empirical result shows that human capital externality has no causal impact on hours worked for the self-employed, but it affects hours worked of employees. This difference comes from the relative strength of the relationship between human capital externalities and localization for the self-employed and employees.

The value and the main task of this paper is the study of the mechanism behind the different patterns. In the literature, there are many studies on the causality running from human capital spillover to agglomeration. This paper also pays attention to the reverse causality, and uses the relationship between human capital externality and agglomeration from both directions as the mechanism to explain the causality between human capital spillover and hours worked. The empirical results show that human capital and urbanization have much stronger connection than human capital and localization. For the self-employed, the relationship between human capital and localization is insignificant. However, only localization has a causal impact on hours worked according to the literature. Therefore, the causality between human capital externality and hours worked of the self-employed is trivial as well. On the other hand, human capital externalities affect localization for employees. Thus, human capital spillovers have impact on hours worked of employees through localization. The robustness check using wages supports our mechanism analysis.

As suggested in Chapter I, direct promotion of the process of localization could also bring benefits to local economies, but this is only for the self-employed. This paper provides some evidence that increasing the local education level could accelerate the agglomeration of employees within industries, and hence increase the size of regional economic activities. Therefore, combining the policies suggested in Chapter I and increasing local education level could generate more benefits for regional economies. To achieve this goal, on the one hand, local governments should consider how to attract people with higher educational backgrounds. On the other hand, policies against brain drain are also necessary.

Although this paper studies the relationship between agglomeration and human capital externalities as the mechanism through which human capital externalities affect hours worked, it is only an empirical work. The underlying intuitions still remain unknown, including why human capital externalities only affect localization for employees but not for the self-employed, and how human capital externalities could affect hours worked through other channels, etc. The future work will address these questions to have a better understanding of the relationship between human capital spillovers and labor market outcomes

REFERENCES

Acs, Z., & Armington, C. (2004). Employment growth and entrepreneurial activity in cities. Regional Studies, 38(8), 911-927.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. Journal of Econometrics, 121(1), 343-375.

Aronson, RobertL. (1991). Self-Employment: A Labor Market Perspective. Ithaca: ILR Press.

Audretsch, D. B., & Keilbach, M. (2004). Does entrepreneurship capital matter? *Entrepreneurship Theory and Practice*, 28(5), 419-429.

Audretsch, D. B., Keilbach, M. C., & Lehmann, E. E. (2006). *Entrepreneurship and economic growth*. New York: Oxford University Press.

Beeson, P. E., & Eberts, R. W. (1989). Identifying productivity and amenity effects in interurban wage differentials. *The Review of Economics and Statistics*, *71*(3), 443-452.

Bernardo, A. E., & Welch, I. (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy*, *10*(3), 301-330.

Berry, C. R., & Glaeser, E. L. (2005). The divergence of human capital levels across cities. *Papers in regional science*, *84*(3), 407-444.

Blanchflower, D. G. (2000). Self-employment in OECD countries. *Labour Economics*, 7(5), 471-505.

Borjas, G. J. (1986). The self-employment experience of immigrants. *Journal of Human Resources*, 21(4), 485–506.

Borjas, G. J., & Bronars, S. G. (1989). Consumer Discrimination and Self-Employment. *The Journal of Political Economy*, *97*(3), 581-605.

Bruce, D., Deskins, J. A., Hill, B. C., & Rork, J. C. (2009). (Small) business activity and state economic growth: Does size matter? *Regional Studies*, 43(2), 229-245.

Brülhart, M., & Sbergami, F. (2009). Agglomeration and growth: Cross-country evidence. *Journal of Urban Economics*, *65*(1), 48-63.

Carlino, G., & Kerr, W. R. (2014). *Agglomeration and innovation* (No. w20367). National Bureau of Economic Research.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review*, *46*(2): 213– 27.

Ciccone, A., and Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, *86* (1): 54– 70.

Combes, P. P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics* (pp. 15-66). University of Chicago Press.

Constant, A., & Zimmermann, K. F. (2006). The making of entrepreneurs in Germany: Are native men and immigrants alike? *Small Business Economics*, 26(3), 279-300.

de Meza, D., & Southey, C. (1996). The borrower's curse: optimism, finance and entrepreneurship. *Economic Journal*, *106*(435), 375-386.

Devine, T. J. (1994). Characteristics of self-employed women in the United States. *Monthly Labor Review*, 20-34.

Djankov, S., Qian, Y., Roland, G., & Zhuravskaya, E. (2006). Who are China's entrepreneurs? *American Economic Review*, 96(2), 348–352.

Duranton, G., & Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *The American Economic Review*, *91*(5), 1454-1477.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J.V., & Thisse, J. F. (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4. (pp. 2063-2117). North-Holland, Amsterdam.

Duranton, G., & Puga, D. (2014). The growth of cities. In Aghion, P., & Durlauf, S. (Eds.), *Handbook of Economic Growth*, Vol. 2. North-Holland, Amsterdam.

Eide, E. R., Hilmer, M. J., & Showalter, M. H. (2015). Is it where you go or what you study? The relative influence of college selectivity and college major on earnings. *Contemporary Economic Policy*, Forthcoming, DOI: 10.1111/coep.12115.

Evans, D. S., & Leighton, L. S. (1989). Some Empirical Aspects of Entrepreneurship. *American Economic Review*, *79*(3), 519-35.

Faggio, G., & Silva, O. (2014). Self-employment and entrepreneurship in urban and rural labour markets. *Journal of Urban Economics*, *84*, 67–85.

Fairlie, R. W. (2006). Entrepreneurship among disadvantaged groups: An analysis of the dynamics of self-employment by gender, race and education. In S. C. Parker, Z. J. Acs, & D. R. Audretsch (Eds.), *Handbook of entrepreneurship, vol. 2* (437−478). Dordrecht, Netherlands: Kluwer Academic Publishers.

Fairlie, R. W., & Meyer, B. D. (1996). Ethnic and racial self-employment differences and possible explanations. *Journal of Human Resources*, 31(4), 757–793.

Fu, S. (2007). Smart café cities: Testing human capital externalities in the Boston metropolitan area. *Journal of Urban Economics*, *61*(1), 86-111.

Glaeser, E. L. (2007). Entrepreneurship and the City (No. w13551). National Bureau of Economic Research.

Glaeser, E. L., and Maré, D. C. (2001). Cities and skills. *Journal of Labor Economics*, *19* (2), 316– 42.

Glaeser, E. L., & Saiz, A. (2003). *The rise of the skilled city* (No. w10191). National Bureau of Economic Research.

Halek, M., & Eisenhauer, J. G. (2001). Demography of risk aversion. *Journal of Risk and Insurance*, 1-24.

Hamilton, B. H. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political economy*, *108*(3), 604-631.

Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2(2), 31-56.

Hurst, E., & Pugsley, B. W. (2011). What do Small Businesses Do? *Brookings Papers on Economic Activity*, *43*(2), 73-142.

Iranzo, S., & Peri, G. (2009). Schooling externalities, technology, and productivity: Theory and Evidence from US states. *The Review of Economics and Statistics*, *91*(2), 420-431.

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available from http://srtm.csi.cgiar.org.

Jianakoplos, N. A., & Bernasek, A. (1998). Are women more risk averse?. *Economic inquiry*, *36*(4), 620.

Kahn, S., & Lang, K. (1991). The Effect of Hours Constraints on Labor Supply Estimates. *The Review of Economics and Statistics*, *73*(4), 605-611.

Kihlstrom, R. E., & Laffont, J. J. (1979). A general equilibrium entrepreneurial theory of firm formation based on risk aversion. *The Journal of Political Economy*, 719-748.

Lassmann, A., & Busch, C. (2015). Revisiting native and immigrant entrepreneurial activity. *Small Business Economics*, Forthcoming, DOI 10.1007/s11187-015-9665-x.

Levine, R., & Rubinstein, Y. (2012). *Does Entrepreneurship Pay? The Michael Bloombergs, the Hot Dog Vendors, and the Returns to Self-Employment*. Working Papers, Haas School of Business, University of California, Berkeley.

Lieberson, S., & Waters, M. C. (1988). From many strands: Ethnic and racial groups in contemporary America. Russell Sage Foundation.

Light, I. (1972). Ethnic Enterprise in America. Berkeley: University of California Press.

Light, I. (1979). Disadvantaged minorities in self-employment. International Journal of Comparative Sociology, 20(1-2), 31-45.

Light, I. (1984). Immigrant and ethnic enterprise in North America. Ethnic and racial studies, 7(2), 195-216.

Liñán, F., & Fernandez-Serrano, J. (2014). National culture, entrepreneurship and economic development: Different patterns across the European Union. *Small Business Economics*, *42*(4), 685-701.

Lofstrom, M. (2004). *Labor market assimilation and the self-employment decision of immigrant entrepreneurs* (pp. 191-222). Berlin Heidelberg: Springer.

Meyer, B. D. (1990). *Why are there so few Black Entrepreneurs?* (No. w3537). National Bureau of Economic Research.

Moore, R. L. (1983). Employer discrimination: Evidence from self-employed workers. *The Review of Economics and Statistics*, 496-501.

Moretti, E. (2004a). Human capital externalities in cities. In Henderson, J.V., & Thisse, J. F. (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4. (pp. 2243-2291). North-Holland, Amsterdam.

Moretti, E. (2004b). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, *121*, 175 – 212.

Peri, G., Shih, K., & Sparber, C. (2015). STEM workers, H-1B visas, and productivity in U.S. cities. *Journal of Labor Economics*, 33(3), S225-S255.

Portes, A., & Zhou, M. (1991). Gaining the Upper Hand: Old and New Perspectives in the Study of the Foreign-Born Minorities. In *Conference paper. Washington, DC: Joint Center for Political Studies*.

Portes, A., & Zhou, M. (1996). Self-employment and the earnings of immigrants. *American Sociological Review*, 219-230.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, *90*(6), 1257-1278.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In Henderson, J.V., & Thisse, J. F. (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4. (pp. 2119-2171). North-Holland, Amsterdam.

Rosenthal, S. S., & Strange, W. C. (2008a). Agglomeration and hours worked. *The Review of Economics and Statistics*, *90*(1), 105-118.

Rosenthal, S. S., & Strange, W. C. (2008b). The attenuation of human capital spillovers. *Journal of Urban Economics*, *64*(2), 373-389.

Ruggles, S., Trent, A. J., Genadek, K., Goeken, R., Schroeder, M. B., & Sobek, M. (2010). *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota.

Sowell, T. (1981). *Markets and minorities*. New York: Basic Books.

United States Department of Justice. Federal Bureau of Investigation. Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 2000. ICPSR03451-v4. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2006-01-16. http://doi.org/10.3886/ICPSR03451.v4

Winters, J. V. (2009). Wages and prices: Are workers fully compensated for cost of living differences? *Regional Science and Urban Economics*, *39*, 632–643.

Winters, J. V. (2013). Human capital externalities and employment differences across metropolitan areas of the USA. *Journal of Economic Geography*, *13*, 799–822.

Winters, J. V. (2014). Foreign and native-born STEM graduates and innovation intensity in the United States. IZA Discussion Paper 8575.

Winters, J. V., & Xu, W. (2014). Geographic differences in the earnings of economics majors. *Journal of Economic Education*, 45(3), 262-276.

Yuengert, A. M. (1995). Testing hypotheses of immigrant self-employment. *Journal of Human Resources*, 30(1), 194–204.

APPENDICES

1 The List of Majors in STEM Fields

Appendix Table 1 lists the majors in STEM fields. The detailed criterion of being selected as STEM is described in Section 3 of Chapter 2.

Appendix Table 1: Majors in STEM Fields

CODE	Major
1103	Animal Sciences
1104	Food Science
1105	Plant Science and Agronomy
1106	Soil Science
1301	Environmental Science
1302	Forestry
2001	Communication Technologies
2100	Computer and Information Systems
2101	Computer Programming and Data Processing
2102	Computer Science
2105	Information Sciences
2106	Computer Information Management and Security
2107	Computer Networking and Telecommunications
2400	General Engineering
2401	Aerospace Engineering
2402	Biological Engineering
2403	Architectural Engineering
2404	Biomedical Engineering
2405	Chemical Engineering
2406	Civil Engineering
2407	Computer Engineering
2408	Electrical Engineering
2409	Engineering Mechanics, Physics, and Science
2410	Environmental Engineering

2411	Geological and Geophysical Engineering
2412	Industrial and Manufacturing Engineering
2413	Materials Engineering and Materials Science
2414	Mechanical Engineering
2415	Metallurgical Engineering
2416	Mining and Mineral Engineering
2417	Naval Architecture and Marine Engineering
2418	Nuclear Engineering
2419	Petroleum Engineering
2499	Miscellaneous Engineering
2500	Engineering Technologies
2501	Engineering and Industrial Management
2502	Electrical Engineering Technology
2503	Industrial Production Technologies
2504	Mechanical Engineering Related Technologies
2599	Miscellaneous Engineering Technologies
3600	Biology
3601	Biochemical Sciences
3602	Botany
3603	Molecular Biology
3604	Ecology
3605	Genetics
3606	Microbiology
3607	Pharmacology
3608	Physiology
3609	Zoology
3611	Neuroscience
3699	Miscellaneous Biology
3700	Mathematics
3701	Applied Mathematics
3702	Statistics and Decision Science
3801	Military Technologies
4002	Nutrition Sciences
4005	Mathematics and Computer Science
4006	Cognitive Science and Biopsychology
5000	Physical Sciences
5001	Astronomy and Astrophysics
5002	Atmospheric Sciences and Meteorology
5003	Chemistry
5004	Geology and Earth Science
5005	Geosciences
5006	Oceanography
5007	Physics
5008	Materials Science
5098	Multi-disciplinary or General Science

5102	Nuclear, Industrial Radiology, and Biological Technologies
5901	Transportation Sciences and Technologies
6106	Health and Medical Preparatory Programs
6108	Pharmacy, Pharmaceutical Sciences, and Administration
6202	Actuarial Science
6212	Management Information Systems and Statistics

2 The Estimates of Detailed Controls for Table 2.2

Appendix Table 2 presents the results for individual control variables of Table 2.2. The estimated coefficients for fixed effects are omitted for space conservation.

Appendix Table 2: Results for Individual Control Variables for Self-Employment Outcomes

	Self-Employment			Incoporated Self-Eployment		
	(1)	(2)	(3)	(4)	(5)	(6)
STEM	-0.0306***	-0.0160***	-0.0153***	-0.0137***	-0.0091***	-0.0087***
	(0.0039)	(0.0034)	(0.0033)	(0.0021)	(0.0023)	(0.0022)
A. Educational Attainments						
Bachelor's Degree	Reference Category					
Master's Degree	-0.0228***	-0.0019	-0.0010	-0.0108***	-0.0010	-0.0004
	(0.0029)	(0.0026)	(0.0026)	(0.0022)	(0.0022)	(0.0021)
Professional Degree	0.0657***	0.0576***	0.0580***	0.0435***	0.0293***	0.0299***
	(0.0070)	(0.0056)	(0.0057)	(0.0056)	(0.0037)	(0.0037)
Doctoral Degree	-0.0350***	0.0153***	0.0167***	-0.0133***	0.0079**	0.0090**
	(0.0056)	(0.0049)	(0.0051)	(0.0034)	(0.0037)	(0.0039)
B. Original Nationality						
Other	Reference Category					
Canada	0.1095***	0.0908***	0.0789**	0.0492***	0.0353**	0.0294
	(0.0151)	(0.0308)	(0.0328)	(0.0137)	(0.0168)	(0.0184)
Mexico	0.0695***	0.0526*	0.0319	0.0151	0.0043	-0.0036
	(0.0181)	(0.0317)	(0.0332)	(0.0188)	(0.0195)	(0.0215)
Rest of Americas	0.0714***	0.0569*	0.0454	0.0223	0.0126	0.0001
	(0.0171)	(0.0306)	(0.0322)	(0.0178)	(0.0188)	(0.0195)
Western Europe	0.0924***	0.0772**	0.0658**	0.0374**	0.0261	0.0203
	(0.0169)	(0.0305)	(0.0323)	(0.0176)	(0.0191)	(0.0206)

Eastern Europe	0.1115***	0.0787***	0.0725**	0.0446***	0.0270	0.0231
	(0.0148)	(0.0299)	(0.0318)	(0.0154)	(0.0174)	(0.0190)
China	0.0681***	0.0612*	0.0520	0.0267*	0.0172	0.0145
	(0.0154)	(0.0315)	(0.0338)	(0.0146)	(0.0178)	(0.0199)
Japan	0.0861***	0.0787**	0.0671*	0.0219	0.0133	0.0095
	(0.0202)	(0.0330)	(0.0349)	(0.0177)	(0.0199)	(0.0218)
Korea	0.1323***	0.1117***	0.1027***	0.0644***	0.0466***	0.0432**
	(0.0117)	(0.0325)	(0.0338)	(0.0145)	(0.0172)	(0.0196)
Philippines	0.0176	0.0273	0.0131	0.0001	-0.0013	-0.0057
	(0.0185)	(0.0323)	(0.0347)	(0.0182)	(0.0200)	(0.0223)
India	0.0802***	0.0654**	0.0599*	0.0407**	0.0256	0.0227
	(0.0170)	(0.0314)	(0.0331)	(0.0163)	(0.0183)	(0.0199)
Rest of Asia	0.1026***	0.0805***	0.0701**	0.0447***	0.0276*	0.0241
	(0.0129)	(0.0306)	(0.0329)	(0.0131)	(0.0165)	(0.0189)
Africa	0.0681***	0.0584*	0.0532	0.0236	0.0151	0.0132
	(0.0196)	(0.0316)	(0.0334)	(0.0182)	(0.0189)	(0.0205)
Oceania	0.0801***	0.0598*	0.0514	0.0305*	0.0184	0.0145
	(0.0180)	(0.0330)	(0.0349)	(0.0173)	(0.0191)	(0.0208)
C. English Speaking						
Does Not Speak	Reference Category					
Speaks Only English	-0.0497***	-0.0004	0.0052	-0.0073	0.0057	0.0121
	(0.0188)	(0.0177)	(0.0184)	(0.0094)	(0.0084)	(0.0092)
Speaks Very Well	-0.0402**	0.0058	0.0103	0.0006	0.0122	0.0176*
	(0.0195)	(0.0189)	(0.0192)	(0.0089)	(0.0087)	(0.0096)
Speaks Well	-0.0284	0.0052	0.0090	-0.0008	0.0064	0.0109
	(0.0182)	(0.0183)	(0.0183)	(0.0083)	(0.0085)	(0.0093)
Not well	-0.0152	-0.0105	-0.0069	-0.0003	-0.0037	-0.0000
	(0.0195)	(0.0184)	(0.0180)	(0.0096)	(0.0089)	(0.0095)
D. Demographic Controls						
Age	-0.0043	-0.0091	-0.0158	-0.0035	-0.0059	-0.0104
	(0.0453)	(0.0414)	(0.0419)	(0.0334)	(0.0318)	(0.0322)
Age Square	0.0001	0.0002	0.0005	-0.0000	0.0001	0.0002
	(0.0017)	(0.0016)	(0.0016)	(0.0013)	(0.0012)	(0.0012)
Age^3	0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Age^4	-0.0000	-0.0000	0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Female	-0.0507***	-0.0378***	-0.0367***	-0.0323***	-0.0231***	-0.0224***
	(0.0053)	(0.0050)	(0.0051)	(0.0034)	(0.0033)	(0.0034)
Married	-0.0043	-0.0002	0.0007	0.0108***	0.0121***	0.0126***
	(0.0051)	(0.0046)	(0.0047)	(0.0032)	(0.0033)	(0.0034)
Married × Female	0.0162***	0.0144***	0.0135***	0.0015	-0.0008	-0.0011
	(0.0048)	(0.0044)	(0.0045)	(0.0034)	(0.0035)	(0.0034)
Citizenship	-0.0027	0.0055	0.0046	0.0061***	0.0082***	0.0071***
	(0.0052)	(0.0040)	(0.0040)	(0.0023)	(0.0021)	(0.0021)
Childen Present	0.0050*	0.0091**	0.0104***	0.0049**	0.0061**	0.0067***
	(0.0030)	(0.0035)	(0.0035)	(0.0023)	(0.0026)	(0.0025)
0 - 5 Years in USA	Reference Category					
6 - 10 Years in USA	0.0169***	0.0064	0.0061	0.0058**	0.0021	0.0019
	(0.0057)	(0.0044)	(0.0044)	(0.0026)	(0.0025)	(0.0026)
11 - 15 Years in USA	0.0276***	0.0176***	0.0177***	0.0136***	0.0099***	0.0106***
	(0.0046)	(0.0043)	(0.0044)	(0.0029)	(0.0028)	(0.0029)
16 - 20 Years in USA	0.0323***	0.0233***	0.0240***	0.0163***	0.0117***	0.0133***
	(0.0062)	(0.0056)	(0.0056)	(0.0044)	(0.0041)	(0.0043)
21+ Years in USA	0.0342***	0.0282***	0.0279***	0.0197***	0.0169***	0.0186***
	(0.0051)	(0.0044)	(0.0044)	(0.0029)	(0.0029)	(0.0035)
Constant	0.0432	0.0538	0.0788	0.0395	0.0626	0.0913
	(0.4432)	(0.4035)	(0.4087)	(0.3308)	(0.3139)	(0.3197)
Industry FE	No	Yes	Yes	No	Yes	Yes
MSA FE	No	No	Yes	No	No	Yes

3 Industry Ranking of Self-Employment Rates

Appendix Table 3 shows the ranking of self-employment and incorporated self-employment rates for all 223 industries in the sample based on the estimated marginal effects. The results come from the preferred specification with both industry fixed effects and MSA fixed effects. Agricultural production, corps is the reference category.

Appendix Table 3: Results for Industry Indicator Variables

Code	Industry	Self-employed Estimates	Self-employed Rank	Incorporated Estimates	Incorporated Rank
893	Miscellaneous professional and related services	0.3383	1	0.0983	10
821	Offices and clinics of chiropractors	0.3075	2	0.3064	1
551	Farm-product raw materials	0.2674	3	-0.1148	223
402	Taxicab service	0.2258	4	0.0469	26
761	Private households	0.2068	5	-0.0452	120
671	Direct selling establishments	0.2015	6	0.0358	32
772	Beauty shops	0.1519	7	0.0570	21
760	Miscellaneous repair services	0.1487	8	-0.0447	118
32	Fishing, hunting, and trapping	0.1465	9	0.0958	11
822	Offices and clinics of optometrists	0.1432	10	0.1073	8
230	Logging	0.1303	11	-0.0653	168
780	Barber shops	0.1189	12	-0.0081	74
820	Offices and clinics of dentists	0.0897	13	0.1701	3
20	Landscape and horticultural services	0.0796	14	0.0286	40
681	Retail florists	0.0769	15	0.0088	57
791	Miscellaneous personal services	0.0695	16	0.0314	37
712		0.0665	17	0.0700	17
	Real estate, including real estate-insurance offices				
611	Food stores, n.e.c.	0.0660	18	0.1262	5
650	Liquor stores	0.0650	19	0.2066	2
750	Automobile parking and carwashes	0.0631	20	0.0122	54
722	Services to dwellings and other buildings	0.0405	21	0.0244	44
682	Miscellaneous retail stores	0.0396	22	0.0904	13
60	All construction	0.0368	23	0.0567	22
800	Theaters and motion pictures	0.0320	24	0.0714	16
860	Educational services, n.e.c.	0.0312	25	0.0063	60
741	Business services, n.e.c.	0.0177	26	0.0337	35
663	Catalog and mail order houses	0.0135	27	0.0296	39
691	Retail trade, n.s.	0.0115	28	0.0344	34
771	Laundry, cleaning, and garment services	0.0110	29	0.0944	12
660	Jewelry stores	0.0069	30	0.0395	31
752	Electrical repair shops	0.0050	31	0.0417	29
10	Agricultural production, crops	0.0000	32	0.0000	65
812	Offices and clinics of physicians	-0.0017	33	0.1165	7
892	Management and public relations services	-0.0090	34	0.0534	24
562	Miscellaneous wholesale, nondurable goods	-0.0208	35	0.1017	9
531	Scrap and waste materials	-0.0215	36	0.1217	6
11	Agricultural production, livestock	-0.0216	37	-0.0202	84
410	Trucking service	-0.0301	38	0.0352	33
862	Child day care services	-0.0333	39	-0.0120	79
142	Yarn, thread, and fabric mills	-0.0367	40	0.0842	14
501	Furniture and home furnishings	-0.0408	41	0.0760	15
280	Other primary metal industries	-0.0457	42	0.1477	4
662	Sewing, needlework, and piece goods stores	-0.0480	43	0.0443	27
500	Motor vehicles and equipment	-0.0520	44	0.0168	51
751	Automotive repair and related services	-0.0525	45	0.0276	42

721	Advertising	-0.0616	46	0.0215	48
601	Grocery stores	-0.0618	47	0.0580	20
621	Gasoline service stations	-0.0673	48	0.0674	19
841	Legal services	-0.0679	49	0.0101	56
542	Apparel, fabrics, and notions	-0.0684	50	0.0231	47
610	Retail bakeries	-0.0688	51	0.0309	38
890	Accounting, auditing, and bookkeeping services	-0.0734	52	0.0106	55
502	Lumber and construction materials	-0.0784	53	0.0527	25
620	Auto and home supply stores	-0.0811	54	0.0423	28
571	Wholesale trade, n.s.	-0.0829	55	0.0068	59
641	Eating and drinking places	-0.0870	56	0.0332	36
582	Retail nurseries and garden stores	-0.0926	57	0.0402	30
810	Miscellaneous entertainment and recreation services	-0.0964	58	-0.0113	78
631	Furniture and home furnishings stores	-0.0971	59	0.0129	53
511	Metals and minerals, except petroleum	-0.1042	60	0.0544	23
180	Plastics, synthetics, and resins	-0.1050	61	-0.0634	155
151	Apparel and accessories, except knit	-0.1079	62	-0.0051	69
861	Job training and vocational rehabilitation services	-0.1151	63	0.0235	46
221	Footwear, except rubber and plastic	-0.1174	64	0.0687	18
731	Personnel supply services	-0.1174	65	-0.0158	81
623	Apparel and accessory stores, except shoe	-0.1245	66	0.0053	62
432	Services incidental to transportation	-0.1257	67	-0.0092	76
661	Gift, novelty, and souvenir shops	-0.1271	68	0.0169	50
633	Radio, TV, and computer stores	-0.1299	69	-0.0030	68
261	Pottery and related products	-0.1325	70	-0.0736	193
882	Engineering, architectural, and surveying services	-0.1327	71	0.0081	58
391	Miscellaneous manufacturing industries	-0.1335	72	0.0000	66
710	Security, commodity brokerage, and investment companies	-0.1337	73	-0.0004	67
172	Printing, publishing, and allied industries, except newspapers	-0.1356	74	-0.0087	75
530	Machinery, equipment, and supplies	-0.1366	75	0.0210	49
532	Miscellaneous wholesale, durable goods	-0.1374	76	0.0243	45
541	Drugs, chemicals, and allied products	-0.1451	77	-0.0092	77
651	Sporting goods, bicycles, and hobby stores	-0.1463	78	0.0281	41
392	Manufacturing industries, n.s.	-0.1501	79	-0.0057	71
732	Computer and data processing services	-0.1517	81	-0.0135	80
12	Veterinary services	-0.1517	80	-0.0162	82
840	Health services, n.e.c.	-0.1525	82	-0.0216	85
612	Motor vehicle dealers	-0.1548	83	-0.0243	91
111	Bakery products	-0.1556	84	0.0061	61
152	Miscellaneous fabricated textile products	-0.1619	85	0.0246	43
171	Newspaper publishing and printing	-0.1637	86	-0.0234	90
652	Book and stationery stores	-0.1648	87	-0.0272	96
112	Sugar and confectionery products	-0.1657	88	0.0015	64
561	Farm supplies	-0.1683	89	0.0137	52

231	Sawmills, planing mills, and millwork	-0.1698	90	0.0030	63
550	Groceries and related products	-0.1716	91	-0.0293	100
581	Hardware stores	-0.1718	92	-0.0051	70
702	Credit agencies, n.e.c.	-0.1734	93	-0.0221	87
580	Lumber and building material retailing	-0.1775	94	-0.0322	102
510	Professional and commercial equipment and supplies	-0.1783	95	-0.0250	94
512	Electrical goods	-0.1786	96	-0.0066	72
871	Social services, n.e.c.	-0.1797	97	-0.0450	119
870	Residential care facilities, without nursing	-0.1805	98	-0.0220	86
122	Food industries, n.s.	-0.1825	99	-0.0067	73
471	Sanitary services	-0.1832	100	-0.0245	92
740	Detective and protective services	-0.1834	101	-0.0373	108
521	Hardware, plumbing and heating supplies	-0.1895	102	-0.0523	133
711	Insurance	-0.1909	104	-0.0459	121
282	Fabricated structural metal products	-0.1909	103	-0.0540	138
101	Dairy products	-0.1934	105	-0.0503	127
140	Dyeing and finishing textiles, except wool and knit goods	-0.1946	106	-0.0230	89
390	Toys, amusement, and sporting goods	-0.1948	107	-0.0384	111
281	Cutlery, handtools, and general hardware	-0.1975	108	-0.0225	88
742	Automotive rental and leasing, without drivers	-0.1976	109	-0.0366	106
951	Coast Guard	-0.1982	110	-0.0632	154
291	Metal forgings and stampings	-0.1985	111	-0.0687	181
242	Furniture and fixtures	-0.1991	112	-0.0490	123
110	Grain mill products	-0.1998	113	-0.0245	93
300	Miscellaneous fabricated metal products	-0.1999	114	-0.0197	83
331	Machinery, except electrical, n.e.c.	-0.2005	115	-0.0381	109
440	Radio and television broadcasting and cable	-0.2011	116	-0.0370	107
762	Hotels and motels	-0.2016	117	-0.0351	103
540	Paper and paper products	-0.2037	118	-0.0250	95
640	Music stores	-0.2042	119	-0.0774	201
642	Drug stores	-0.2072	120	-0.0381	110
630	Shoe stores	-0.2076	121	-0.0443	116
560	Alcoholic beverages	-0.2100	122	-0.0287	99
781	Funeral service and crematories	-0.2102	123	-0.0709	189
372	Medical, dental, and optical instruments and supplies	-0.2106	124	-0.0361	104
120	Beverage industries	-0.2107	125	-0.0315	101
872	Museums, art galleries, and zoos	-0.2108	126	-0.0503	128
370	Cycles and miscellaneous transportation equipment	-0.2110	127	-0.0275	97
340	Household appliances	-0.2114	128	-0.0285	98
100	Meat products	-0.2127	129	-0.0565	143
891	Research, development, and testing services	-0.2146	130	-0.0496	125
121	Misc. food preparations and kindred products	-0.2153	131	-0.0364	105
161	Miscellaneous paper and pulp products	-0.2174	132	-0.0443	117
552	Petroleum products	-0.2184	133	-0.0472	122
212	Miscellaneous plastics products	-0.2187	134	-0.0616	152

272	Primary aluminum industries	-0.2202	135	-0.0396	112
401	Bus service and urban transit	-0.2205	136	-0.0409	114
42	Oil and gas extraction	-0.2216	137	-0.0576	144
371	Scientific and controlling instruments	-0.2239	138	-0.0442	115
622	Miscellaneous vehicle dealers	-0.2239	139	-0.0750	198
832	Nursing and personal care facilities	-0.2247	141	-0.0517	132
210	Tires and inner tubes	-0.2247	140	-0.0648	167
801	Video tape rental	-0.2280	142	-0.0514	131
591	Department stores	-0.2284	143	-0.0581	145
600	Miscellaneous general merchandise stores	-0.2284	144	-0.0594	146
852	Libraries	-0.2291	145	-0.0661	173
312	Construction and material handling machines	-0.2295	146	-0.0510	130
270	Blast furnaces, steelworks, rolling and finishing mills	-0.2296	147	-0.0508	129
700	Banking	-0.2296	148	-0.0602	149
360	Ship and boat building and repairing	-0.2306	149	-0.0530	136
351	Motor vehicles and motor vehicle equipment	-0.2307	151	-0.0550	141
182	Soaps and cosmetics	-0.2307	150	-0.0554	142
292	Ordnance	-0.2319	152	-0.0757	200
332	Machinery, n.s.	-0.2321	153	-0.0530	137
670	Vending machine operators	-0.2326	155	-0.0547	140
361	Railroad locomotives and equipment	-0.2326	154	-0.0647	165
192	Industrial and miscellaneous chemicals	-0.2331	156	-0.0526	134
262	Misc. nonmetallic mineral and stone products	-0.2357	157	-0.0526	135
420	Water transportation	-0.2358	158	-0.0634	156
251	Cement, concrete, gypsum, and plaster products	-0.2360	159	-0.0494	124
441	Telephone communications	-0.2362	160	-0.0637	157
942	Navy	-0.2379	161	-0.0596	147
341	Radio, TV, and communication equipment	-0.2380	162	-0.0681	177
342	Electrical machinery, equipment, and supplies, n.e.c.	-0.2383	163	-0.0601	148
271	Iron and steel foundries	-0.2388	164	-0.0662	174
842	Elementary and secondary schools	-0.2399	165	-0.0641	161
250	Glass and glass products	-0.2401	166	-0.0399	113
940	Army	-0.2401	167	-0.0644	163
352	Aircraft and parts	-0.2408	168	-0.0541	139
421	Air transportation	-0.2410	169	-0.0656	171
181	Drugs	-0.2411	170	-0.0677	176
222	Leather products, except footwear	-0.2421	171	-0.0502	126
850	Colleges and universities	-0.2421	172	-0.0647	166
252	Structural clay products	-0.2431	173	-0.0640	159
831	Hospitals	-0.2449	175	-0.0641	162
41	Coal mining	-0.2449	174	-0.0667	175
881	Membership organizations, n.e.c.	-0.2453	176	-0.0617	153
211	Other rubber products, and plastics footwear and belting	-0.2457	178	-0.0610	151
162	Paperboard containers and boxes	-0.2457	177	-0.0640	160
400	Railroads	-0.2459	179	-0.0660	172

310	Engines and turbines	-0.2462	180	-0.0783	204
320	Metalworking machinery	-0.2465	181	-0.0639	158
322	Computers and related equipment	-0.2467	182	-0.0654	169
950	Marines	-0.2493	183	-0.0783	205
941	Air Force	-0.2495	184	-0.0645	164
200	Petroleum refining	-0.2504	185	-0.0684	179
873	Labor unions	-0.2505	186	-0.0606	150
190	Paints, varnishes, and related products	-0.2510	187	-0.0694	184
102	Canned, frozen, and preserved fruits and vegetables	-0.2514	188	-0.0684	180
450	Electric light and power	-0.2522	189	-0.0696	186
701	Savings institutions, including credit unions	-0.2523	190	-0.0689	182
932	National security and international affairs	-0.2534	191	-0.0683	178
311	Farm machinery and equipment	-0.2544	192	-0.0861	216
452	Electric and gas, and other combinations	-0.2547	193	-0.0704	188
901	General government, n.e.c.	-0.2551	194	-0.0692	183
472	Utilities, n.s.	-0.2562	195	-0.0736	194
201	Miscellaneous petroleum and coal products	-0.2581	196	-0.0834	211
411	Warehousing and storage	-0.2604	197	-0.0720	192
770	Lodging places, except hotels and motels	-0.2604	198	-0.0745	196
922	Administration of human resources programs	-0.2611	200	-0.0737	195
802	Bowling centers	-0.2611	199	-0.1076	222
40	Metal mining	-0.2622	201	-0.0778	202
952	Armed Forces, branch not specified	-0.2626	202	-0.0798	207
191	Agricultural chemicals	-0.2636	203	-0.0694	185
930	Administration of environmental quality and housing programs	-0.2641	204	-0.0754	199
910	Justice, public order, and safety	-0.2649	205	-0.0781	203
160	Pulp, paper, and paperboard mills	-0.2664	206	-0.0748	197
632	Household appliance stores	-0.2687	207	-0.0703	187
672	Fuel dealers	-0.2696	208	-0.0710	190
960	Military Reserves or National Guard	-0.2696	209	-0.0871	217
301	Metal industries, n.s.	-0.2697	210	-0.0654	170
921	Public finance, taxation, and monetary policy	-0.2697	211	-0.0795	206
931	Administration of economic programs	-0.2699	212	-0.0813	208
130	Tobacco manufactures	-0.2700	213	-0.0909	219
451	Gas and steam supply systems	-0.2706	214	-0.0831	210
50	Nonmetallic mining and quarrying, except fuels	-0.2711	215	-0.0920	220
362	Guided missiles, space vehicles, and parts	-0.2714	216	-0.0827	209
470	Water supply and irrigation	-0.2716	217	-0.0839	212
141	Carpets and rugs	-0.2736	218	-0.0716	191
232	Wood buildings and mobile homes	-0.2736	219	-0.0951	221
132	Knitting mills	-0.2770	220	-0.0860	215
412	U.S. Postal Service	-0.2778	221	-0.0857	214
880	Religious organizations	-0.2803	222	-0.0839	213
31	Forestry	-0.2871	223	-0.0878	218

VITA

Zhengyu Cai

Candidate for the Degree of

Doctor of Philosophy

Thesis: THREE ESSAYS IN APPLIED ECONOMICS ON LABOR SUPPLY DECISIONS

Major Field: Economics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Economics at Oklahoma State University, Stillwater, Oklahoma in May, 2016.

Completed the requirements for the Bachelor of Economics at Southwest Jiaotong University, Chengdu, China in 2010.

Experience:

Research Assistant, Oklahoma State University, 2015-2016.

Instructor, Oklahoma State University, 2014-2015.

Graduate Research Associate, Oklahoma State University, 2013, 2015.

Graduate Teaching Associate, Oklahoma State University, 2011-2012.