Essays on the Labor Market Effects of Globalization

Yubo Liu

Wuhan, China

B.S. in Economics, University of Minnesota, 2012

M.A. in Economics, University of Virginia, 2013

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Economics

University of Virginia

May, 2018

Essays on the Labor Market Effects of Globalization

Yubo Liu

Abstract

The first chapter presents a series of novel facts about US internal migration based on gravity estimations using detailed county migration data from the Internal Revenue Service. To account for the existence of zero flows in the data, I perform Poisson Pseudo Maximum Likelihood and Tobit estimations in addition to Ordinary Least Squares. I find that migrant flows and incomes are positively related to total incomes in the origin and the destination, and negatively related to the distance between those. Households move from areas with lower average incomes, higher housing costs and unemployment rates to areas that are the opposite. The pull effect of average income on migration is stronger than the push effect, while the push effects of housing cost and unemployment rate are stronger than the pull effects. Additionally, average migrant income increases with average incomes in both the origin and the destination, as well as the distance traveled, and decreases with the share of migrants out of an area. These empirical patterns suggest positive selection of migrants.

The second chapter develops a spatial equilibrium model that accounts for the observed migration patterns, and describes the channels through which import competition affects migration. I employ two-stage least squares estimation to identify the causal effects of rising import exposure on migration, exploiting exogenous variations in lowering trade costs and rising Chinese productivities. I find a \$1,000 increase in imports per worker lowers the population by 5.7 percent, while raising the out-migration rate by 1.7 percentage

points in a local labor market. The same increase in imports per worker in the origin raises net migrant outflows by 0.574 log points, and that in the destination lowers net inflows by 0.418 log points. The stronger effect of import exposure in the origin than in the destination implies possible information asymmetry faced by potential migrants who are more familiar with their current residences than future ones.

The third chapter examines the productivity and substitution effects of immigration and offshoring in service sectors in the US. I explore the skill distributions of natives, immigrants and offshore workers to gauge their substitutability. In order to jointly analyze the effects of immigration and offshoring, I match data from the IPUMS samples of the Census-American Community Survey with the multinationals dataset and service imports data from the Bureau of Economic Analysis. I find evidence of substitution between native and offshore workers, and a negative productivity effect of offshoring. Immigration generates positive productivity effect, and immigrants complement native employment. Additionally, the skill complexities of tasks performed by native and immigrant workers increase with more offshoring, and the gap between native and immigrant complexity narrows.

Acknowledgements

I am deeply indebted to my advisors James Harrigan and John McLaren for their continuous guidance and support. I am also grateful for the advice offered by other faculty members at the Department of Economics and the Darden School of Business including Kerem Cosar, Jonathan Colmer, Peter Debaere, Leora Friedberg and William Johnson.

I would like to thank my fellow graduate students Myunghuan Yoo, Jining Zhong, Pingyu He, Ben Hamilton, Hundanol Kebede, Devaki Ghose and Abiy Teshome for helpful comments and suggestions at international trade workshops.

This dissertation couldn't have been completed without the love and support from my parents and my wife, whom I dedicate this work to.

Contents

List of Figures

List of Tables

Chapter 1

Who Moves Where? An Analysis on the Determinants of Internal Migration Patterns in the United States

1.1 Introduction

Since China's accession to the World Trade Organization in 2001, US goods imports from China have grown by 353%, and China's share in US total goods imports has risen from 9% to 22%.[1] Despite a growing body of literature that examines the labor market consequences of intensifying import competition from China, relatively few papers have studied the mobility response to it. The extent to which import competition affects internal migration not only suggests possible attenuation biases in wage and employment outcomes in areas that are directly affected, but sheds light on indirect effects through the migration channel. Therefore, analyses on the rate of geographic labor adjustment are crucial for an accurate assessment of the labor market effects both on the regional and the national level.

[1] As shown in Figure 1.1. Data comes from the Office of the US Trade Representative and the World Bank.

To gain a full understanding of the migration response to import competition, I present a series of novel empirical facts on US internal migration patterns in this chapter. The empirical facts are obtained from estimating a set of gravity equations on internal migration flows under different econometric specifications. In addition to Ordinary Least Squares (OLS), I perform Poisson Pseudo Maximum Likelihood (PPML) and Tobit estimations to account for zero flows and censoring in the data. The empirical strategies take advantage of the largest and most disaggregated migration data from the Internal Revenue Service (IRS) that has rarely been exploited by existing research on internal migration in the US. This data records annual migration flows and incomes between every county-pair in the United States from 1990 to 2015. Data on the county level is then aggregated to the level of Commuting Zones (CZs) that better approximate local labor markets.[2]

I make several observations from gravity estimations on US internal migration flows.[3] First, gross migration flow and income are positively related to total incomes in the origin and the destination, and negatively related to the distance between those locations.[4] Second, households move from areas with lower average incomes, higher housing costs and higher unemployment rates to areas with higher average incomes, lower housing costs and lower unemployment rates. Third, average income in the *destination* has a larger effect on migration flows than that in the *origin*, while housing

[2]Commuting Zones are clusters of counties within which a significant number of people commute to work across county boundaries relative to the population, thus more closely reflect local economies where people live and work. There are 741 CZs in the United States according to the 1990 delineation used in this paper.

[3]Originally developed for estimating international trade flows, the gravity equation has been successful in predicting the empirical relationship between trade flows, importer/exporter GDP and distance between trading partners. It has been widely used in other topics such as international migration, foreign direct investment, etc.

[4]Estimates on total origin and destination incomes are smaller than those in the trade literature, while estimates on distance are bigger in magnitude. From 2,508 estimates in 159 papers that estimate gravity models on trade flows, Head and Mayer, 2014 document the median estimate on origin GDP to be 0.97, the median estimate on destination GDP to be 0.85, and the median estimate on distance to be -0.89. Estimates on origin and destination incomes for US internal migration flows in this paper are less than 0.5, while those on distance are less than -1.2.

price and unemployment rate in the *origin* have larger effects on migration flows than those in the *destination*.[5] Fourth, areas with higher average incomes, higher housing costs, and lower unemployment rates send out richer migrants, and richer migrants move to areas with higher average incomes, lower housing costs and lower unemployment rates. Fifth, average migrant income decreases with migrant share, and increases with the distance traveled. The correlation between average migrant income and migrant share becomes weaker when controlling for distance.

This paper adds to the existing literature by providing the first gravity estimates on US internal migration using the IRS migration data. Owing to its ability to explain how goods, capital and people move between locations, the gravity equation has seen broad applications in previous work since Tinbergen (1962) used it to estimate international trade flows.[6] Portes and Rey (2005) find the gravity equation to explain cross border portfolio patterns as well as they explain trade flows. Head and Ries (2008) draw a similar conclusion for Foreign Direct Investment. On the migration front, Ramos and Surinach (2016) find the distance elasticity to be around -1 from gravity estimations on intra-EU and EU-ENC (EU Neighboring Countries) migration using log immigration stock as the dependent variable. In contrast, Bryan and Morten (2017) use log share of migrants as the dependent variable and estimate distance elasticities of cross-provincial migration in Indonesia and inter-state migration in the US to be around -0.6. It is worth noting that these gravity estimates on migration rely exclusively on the "stock" of migrants at

[5]In other words, the pull effect is stronger than the push effect for average income, while the opposite is true for housing price and unemployment rate.

[6]Trade economists have since provided theoretical foundations for the gravity equation in various settings. Anderson (1979) presented a model based on CES preferences and differentiated goods by the origin country. Anderson and Wincoop (2003) developed a theoretical gravity model to study the border effect in trade flows. Eaton and Kortum (2002) derived and estimated a gravity equation driven by Ricardian comparative advantages.

one point in time, which does not account for return migration and migration to third locations that shape observed patterns.[7] Furthermore, gravity estimates for internal migration in the US do not exploit variations in migration within states and every year. According to the IRS county migration data, 24% of migration flows occur within states, and 44% of migrants move within states between 1992 and 2007. This suggests the average migration flow within states is bigger than that across states. In addition to not exploiting variations in migration within states, estimates on state-to-state flows rely on the assumption that people travel the same distances between the same pair of states, whereas in reality distances between the same pair of states may vary a lot.[8] This paper improves on these fronts by using annual migration data between the most disaggregated geographic units that cover the entire United States.

The remainder of this paper is organized as follows. Section 1.2 describes the data source and provides descriptive evidence from the data. Section 1.3 presents empirical facts on internal migration based on gravity estimations. Section 1.4 concludes and discusses areas for further exploration.

1.2 Data and Descriptive Evidence

1.2.1 Migration Data

Migration flows between pairs of CZs are compiled and aggregated from publicly available data on county migration and incomes through the IRS. The IRS calculates migration flows between two counties by comparing zip codes of tax returns filed by the same household in two consecutive years.

[7] There are time gaps between observed immigrant/migrant stocks due to the structure of population censuses.

[8] For example, the distance between Phoenix, AZ and San Diego, CA is 355 miles, and the distance between Phoenix, AZ and San Francisco, CA is 753 miles. Although the latter distance is more than double the former one, state-to-state flows data would treat them the same.

Both the number of returns (which approximate households) and the number of exemptions (which approximate individuals) are provided for measuring migration. In addition to the number of migrants, the IRS also reports their gross adjusted income from tax returns. Since 87% of households filed tax returns between 1992 and 2007, this is the largest and the most comprehensive dataset for analyzing annual county migration patterns. However, most existing studies rely on migration data from the Census and the March CPS, which suffer from two main disadvantages over the IRS migration data:

First, the Census records the current and previous state or metropolitan area an individual resides every 5 years, and the March CPS records the current and previous state of residence every year. Migration flows between states omit a considerable amount of information on intra-state migration, while migration flows between metro areas omit information on migration flows outside of those areas. In comparison, the IRS county migration data provides migration flows between the smallest geographic units and covers the entire country. Second, calculation of sample weights in the Census and CPS does not account for the share of households or individuals who migrated from a certain area, therefore it may not accurately represent the population based on their previous residence, which is essential for analyses on migration flows between pairs of locations.[9] In contrast, the IRS county migration data comes from the universe of tax filers, thus it is unlikely to be subject to sampling or measurement errors. Third, the smallest geographic division available in the Census is the PUMA (Public Use Microdata Area) or the Super-PUMA, depending on the year of the sample. PUMAs are geographic areas with at least 100,000 residents, and Super-PUMAs are areas

[9]For example, consider three locations A, B, and C, for which the distance between A and B is smaller than the distance between A and C. Suppose we are only interested in estimating the distance elasticity using migration flows from A to B and from A to C. The sample may yield the same proportion of households moving from A to B and from A to C, whereas in reality more households are moving from A to B than from A to C. The distance elasticity estimate from the sample without proper weights will be biased in this case.

with at least 400,000 residents. Since there is no direct matching from PUMAs or Super-PUMAs to Commuting Zones, CZ populations need to be imputed from the Census, which may lead to imprecise estimates.

Despite the advantages over Census and CPS, the IRS migration data does not provide any information on the individual level, though average migrant income can be calculated from total migrant income and migrant flow. Additionally, the IRS data censors migration flows less than 10 households to preserve confidentiality, and may underestimate migration of the poor and the elderly who are less likely to file tax returns.[10] Nevertheless, the fraction of non-filers remained stable over the sample period, which ameliorates concerns for understating the effect of labor demand shocks on migration due to attrition.

Migration Flows and CZ Population

In Table 1.1, I provide summary statistics for annual migration flows (Panel A) and CZ populations (Panel B) from 1992 to 2007. For each row in Panel A, migration takes place between the year in column 1 and the subsequent year. According to Panel A, the median migration flow slightly increased from 29 households in 1992 to 32 households in 2006. There was a more notable increase in the average household flow from 117 to 136 in this period. This implies that distribution of annual migration flows is skewed to the right, and the growth of average migration flows over time is mostly driven by CZ pairs with bigger flows than the median. Panel B shows summary statistics of the population distribution out of 741 CZs each year. The median CZ population grew from 35,615 to 42,127 households, amounting to a 33.4% increase over the sample period. Similar to migration flows, the annual distribution

[10]Migration studies have pointed out that the poor and the elderly are also less likely to migrate.

of CZ population is also skewed to the right, given the disparity between the mean and median household number each year.

Figure 1.2 shows the median household migration flows within and between Census regions. Intra-regional median flows are bigger than inter-regional flows, except for the South which sends more migrants to the Northeast than itself. Among intra-regional flows, Northeast-Northeast is much bigger than the others. These differential patterns in migration are to be accounted for in empirical estimations.

Average Migrant Income and CZ Income

Table 1.2 lists the summary statistics of average incomes for migrants (Panel A) and CZ residents (Panel B). According to Panel A, the median average migrant income grew steadily from $33,815.37 in 1992 to $41,726.31 in 2000, and trended downwards to $38,928.48 in 2003 before picking up again. Similar patterns can be found in the evolution of mean average migrant income (among migration flows) as well as the median and mean CZ income shown in Panel B. This is likely due to the recession between 2000 and 2002.

Similar to that of migration flows, the distribution of average migrant income is also skewed to the right each year, with a large gap between the minimum and maximum. In comparison, the distribution of CZ average income is less skewed and spread out. It is worth noting that average migrant incomes at or below the 50th percentile are consistently smaller than their counterparts in the CZ income distribution, while average migrant incomes at 90th percentile or above are higher. Thus there are more migrant flows with less than the median CZ income than there are above, while flows with more than the median CZ income have very high average income. This points to a large variance in the distribution of migrant incomes.

1.2.2 Trade and Employment Data

I obtain HS 6-digit level trade data from the UN Comtrade database. To construct the import exposure variable, I aggregate annual US imports from China from 1991 to 2007 into 10 main manufacturing industries. I then use the industry employment share for each CZ to apportion variations in industry imports to a specific area, and sum over all industries to obtain import exposure for the entire area. I follow a similar procedure in constructing the instrumental variable, using import data on eight other high-income countries.[11] In Figure 1.3, I compare import growths (from China) in 10 main manufacturing industries between the US and other high-income countries over 1991-2001 and 2001-2007 respectively.[12]

Import growths in US and other developed countries are faster in the first period than the second. The cross-industry variation in growth rates are similar between the US and other developed countries in each period. This similarity suggests that changes in US industry imports are likely driven by factors that also drive import growths in other developed countries, rather than demand shocks specific to the US. The instrumental variable strategy makes use of this empirical pattern, and uses imports of other developed countries in the construction of the instrument for import exposure in the US. The corresponding decrease in the average tariff facing Chinese exports is 83 percent over the entire period.[13]

Industry employment data on the CZ level is extracted from the County Business Patterns (CBP) Database. The only source for complete and consistent annual county-level data on US business establishments by industry,

[11] The eight countries are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland, as those countries have disaggregated data available from 1991.

[12] I choose two periods separated by the year 2001 as it is the year China joined the WTO.

[13] The decrease in tariff is calculated from the World Bank's World Development Indicator. The tariff rate is the weighted mean applied tariff, which is the average of effectively applied rates weighted by the product import shares corresponding to each partner country.

the CBP provides information on the number of establishments, range of employment, etc. that I use to calculate employment shares using a fixed point algorithm developed by Autor, Dorn, and Hanson, 2013. Figure 1.4 shows distributions of import exposure per worker and out-migration rate across CZs in 1992 and 2006. There is a noticeable shift to the right for both distributions, and the shift for import exposure is more dramatic than for the out-migration rate. Figure 1.5 compares import exposure per worker with population growth in commuting zones. According to this figure, the “rust belt" areas experienced greater import exposure and slower population growth from 1991 to 2007.[14] Areas that were not as exposed to import competition, such as southern Texas, Florida and many western states experienced faster population growth.

1.2.3 Data on Local Economic Conditions

I extract data on population, income, housing price and unemployment rate from various sources to create variables that indicate local economic conditions. Data on county population and income (aggregated to the CZ level) comes from the IRS county income database, which provides information on the number of households/individuals and gross adjusted income of tax filers for each county from 1989 to 2015. I use House Price Index (HPI) from the Federal Housing Finance Agency, which measures the average price changes in repeat sales or refinancings on the same properties over time. Since expenditures on housing comprise a third of total expenditures in the US , the HPI also serves as an indicator of overall price changes.[15] Besides wages and prices, unemployment rate is also a potential factor in determining the “attractiveness” of a local area. I use the Local Area Unemployment Statistics

[14]The "Rust Belt" begins in western New York and traverses west through Pennsylvania, West Virginia, Ohio, Indiana, and the Lower Peninsula of Michigan, ending in northern Illinois, eastern Iowa, and southeastern Wisconsin.

[15]Expenditure shares come from the Bureau of Labor Statistics.

(LAUS) from the Bureau of Labor Statistics to calculate unemployment rates. Data on the distance between CZs comes from the NBER County Distance database.

1.3 Initial Observations on US internal migration patterns

In this section, I present a series of empirical observations from gravity estimations. I begin with estimates from the baseline gravity equation, which associates migration flows and incomes with total incomes in the origin and the destination as well as the distance between those. I then break down total incomes into average incomes and populations, which I regress migration flows on, to further investigate the determinants of migration flows. I include variables on housing costs and unemployment rates in the equation to account for factors such as living cost and likelihood of obtaining employment that influence decisions to migrate. Finally, I explore the possibility of selection in migrants by regressing average migrant income on migrant share and distance. The empirical patterns can be summarized as follows:

1. Gross migration flows are positively correlated with total incomes in both the origin and destination, and negatively correlated with the distance between those;

2. Households move from poorer areas to richer areas, and from areas with higher housing costs and unemployment rates to areas with lower housing costs and unemployment rates;

3. The pull effect of average income is stronger than the push effect, while the push effects of housing cost and unemployment rate are stronger than the pull effects;

4. Areas with higher average incomes, higher housing costs and lower unemployment rates send out richer migrants, and richer migrants move to areas with higher average incomes, lower housing costs and lower unemployment rates;

5. Higher average migrant income is associated with a smaller migrant share and a greater distance traveled. The correlation between average migrant income and migrant share becomes weaker when controlling for distance.

1.3.1 Baseline Gravity

First, I estimate a gravity equation on US internal migration flows that resembles the gravity equation in trade, which relates international trade flows to the Gross Domestic Products of the origin and the destination country, and the distance between those countries. The estimating equation for migration flows is as follows:

$$Y_{dot} = \alpha_0 + \beta_1 I_{ot} + \beta_2 I_{dt} + \beta_3 Dist_{do} + \gamma_o + \gamma_d + \gamma_{rp} + \gamma_t + \epsilon_{dot}$$

In this equation, o and d denote the origin and the destination CZ, and t denotes year. The dependent variable Y_{dot} is the gross migration flow or the total income of migrants from o to d between t and $t+1$. I_{ot} and I_{dt} stand for total incomes in the origin and the destination. $Dist_{do}$ measures the geographic distance between CZs. All variables are measured in logs. The time fixed effect γ_t captures macroeconomic shocks common to all CZs in a given year. γ_o, the origin fixed effect controls for unobserved common location preferences among migrants from the same origin, where as γ_d, the destination fixed effect captures the "attractiveness" of a destination, such as infrastructure and the quality of education. Additionally, I include the region-pair fixed effect γ_{rp} to control for observed differences in migration between region pairs, as people from the same region may prefer to move within the same region or another region that has a similar climate, etc.

β_1 and β_2 are identified with both cross-sectional and time-series variations in incomes and migration flows.[16] The identification of β_3 relies on the geographic variation of the distances between pairs of CZs. To account for the existence of zero flows, I perform Poisson Pseudo Maximum Likelihood (PPML) and Tobit estimations in addition to Ordinary Least Squares.[17] Similar to gravity estimates for trade, I expect β_1 and β_2 to be positive, and β_3 to be negative.

Results for Migration Flow

Results using the number of migrants as the dependent variable are shown in Table 1.3. OLS estimates are reported in the first four columns, which differ by the fixed effects included. PPML and Tobit estimates are shown in the next four columns. Results in column 1 for PPML and Tobit use the same sample of all positive flows as the OLS for comparison, while those in column 2 use the full sample including zero flows. Total incomes in both the origin and destination CZs are positively correlated with migrant flows, and the coefficient on the destination income is higher than that on the origin income. The distance elasticity falls between -1.3 and -2.1 depending on the estimation method. PPML and Tobit estimates of the distance elasticity using the full sample are greater in magnitude than the OLS estimates, implying a significant amount of zero flows that result from long distances, though

[16] Since migrants comprise a relatively small share of the CZ population, it is reasonable to treat incomes as exogenous to migration flows.

[17] Silva J. and Tenreyro (2006) developed the PPML method to accommodate zero flows by estimating trade flows in levels instead of logs.

the differences are relatively small.[18] Kurzendoerfer (2015) draws a similar conclusion comparing PPML, Tobit and OLS estimates on international trade flows. Coefficient estimates on total incomes are smaller than those on GDP in the trade literature, while distance elasticity estimates are bigger in magnitude than corresponding estimates on trade flows. This suggests greater relative importance of distance versus total incomes in determining US internal migration flows than international trade flows.

Results for Total Migrant Income

Table 1.4 shows estimates using total migrant income as the dependent variable. Coefficient estimates on the origin and destination incomes are both larger than those using the number of migrants as the dependent variable (as shown in Table 3). Therefore incomes in the origin and the destination better predict the total income of migrants than the number of migrants. Distance elasticity estimates are almost identical to those in Table 1.3. Using annual sub-samples to estimate the distance elasticity, I find that it has been steadily increasing over time. This implies rising importance of distance in determining migration flows and declining mobility over this period.

1.3.2 Augmented Gravity

Besides aggregate economic indicators such as the total income of all households in a CZ, variables that measure the average standard of living are potential determinants of migration flows, and entail more information on the local economy than total income. Below I present estimates on augmented gravity models of migration using comprehensive information on local economic conditions and discuss how those conditions relate to gross and net migration flows, as well as average migrant income:

[18] PPML and Tobit estimates using the same sample as OLS are very similar to OLS estimates.

$$Y_{dot} = \alpha_0 + \beta_1 \bar{I}_{ot} + \beta_2 P_{ot} + \beta_3 \bar{I}_{dt} + \beta_4 P_{dt} + \beta_5 HPI_{ot} + \beta_6 UR_{ot} + \beta_7 HPI_{dt}$$
$$+ \beta_8 UR_{dt} + \beta_9 Dist_{do} + \gamma_o + \gamma_d + \gamma_{rp} + \gamma_t + \epsilon_{dot}$$

$\bar{I}_{xt}$ and P_{xt}, where $x \in \{o, d\}$, denote the average income and the population in the origin/destination. HPI_{xt} and UR_{xt} refer to the House Price Index and Unemployment Rate in the origin/destination. The House Price Index proxies for the average cost of living, while the Unemployment Rate indicates the likelihood of gaining employment. Although distance is included, its coefficient estimate is expected to be around zero, as the distance effects in opposite directions of migration flows between the same CZ pair cancel out each other. All variables are in logs except for unemployment rate.

Results for Gross Migration Flow

Table 1.5 shows estimates for gross migration flows. Results using the preferred specification are shown in column 1. Coefficient estimates on both origin and destination population are positive. The origin average income, despite having a negative estimate, does not appear to have any predictive power on migration flows. In comparison, the destination average income is positively correlated with migration flows — a 1% increase in the destination income corresponds to a 0.5% increase in migration flows. Estimates on the origin HPI and UR are positive and significant, while those on the destination HPI and UR are negative and significant. Therefore, migrants move from CZs with higher housing prices and unemployment rates to others that are the opposite. The distance elasticity is similar to previous estimates at around -1.3.

It is worth noting that the HPI and UR in the origin have stronger effects on gross migration than those in the destination. For example, a 1 percentage point increase in the origin UR raises migration by 2.6 percent, whereas the same percentage increase in the destination UR lowers migration only by 1.6 percent. However, for average income, the effect is bigger in the destination than in the origin. This suggests stronger pull effect of average income than push effect, and stronger push effects of housing price and unemployment rate than pull effects.

Results for Net Migration Flow

Results using net migration flows as the dependent variable are shown in Table 1.6 and 1.7. Since a third of CZ pairs each year have recorded migration flows only in one direction, I compare estimates using the full sample in Table 1.6, with estimates using the sample with positive flows in both directions in Table 1.7. In both tables, net flows increase with destination income, while decrease with origin income and population. Therefore, migrants are on net moving from poorer areas to richer areas. Net flows are positively correlated with HPI and UR in the origin, while negatively correlated with those in the destination. This means migrants move from areas with higher living costs and unemployment rates to others that have more favorable living conditions. Although qualitatively the same as results in Table 1.7, estimates in Table 1.6 where all CZ pairs are included are bigger in magnitude, due to possibly widened gaps in bilateral flows when the flow in one direction is censored.

Results for Average Migrant Income

Table 1.8 lists out coefficient estimates using average migrant income as the dependent variable. Average incomes in both the origin and the destination are positively associated with average migrant income – a 1 percent increase in either the origin or the destination average income corresponds to a 0.7 percent increase in the average migrant income. Average migrant income is positively related to distance, meaning richer households migrate to further destinations. Results on HPI and UR suggest that areas with higher housing costs and lower unemployment rates send out richer migrants, who move to areas with lower housing costs and lower unemployment rates.

Table 1.9 shows estimates using the share of migrants out of an origin as the explanatory variable, in addition to other controls. The estimate on migrant share in column 1 is negative and significant, implying higher migrant incomes are associated with smaller shares out of the origin, controlling for average incomes, populations, etc. The correlation weakens when distance is included in the equation, as is shown in column 2. This implies that the distance effect on average migrant income operates through the share of migrants. In other words, greater distances reduce the number of migrants out of an area, and migrants who move long distances are those earning higher incomes. These empirical patterns suggest that migrants are positively selected in terms of their average incomes.

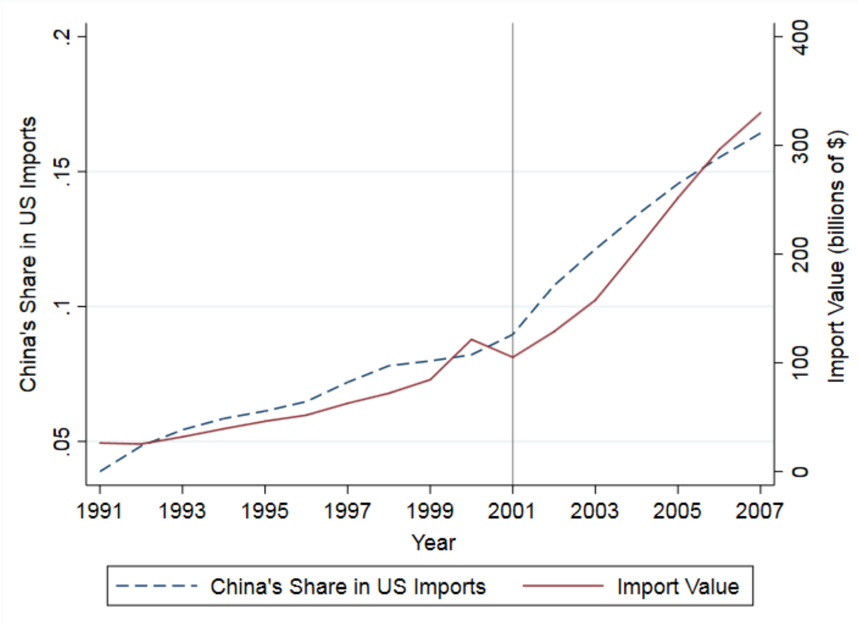
1.4 Conclusion

This chapter has studied the determinants of internal migration patterns in the US and provided the first gravity estimates on US migration using the annual disaggregated IRS migration dataset. In particular, I examine how local economic conditions such as income, housing price and unemployment rate affect migration flows between a pair of locations. Results show that households migrate from areas with worse economic conditions to areas with better economic conditions. In addition, migrants are positively selected in terms of their average income. Similar to gravity estimates on international trade flows, the distance elasticity is negative and significant.

The empirical patterns call for a theoretical framework that places assumptions on the distribution of worker productivities and preferences, and generates empirical predictions that motivate gravity estimations in this chapter. The theoretical framework will be discussed in detail in the next chapter. One area worth further exploration is, upon data availability, using individual-level information to analyze how demographics are associated with migration outcomes.

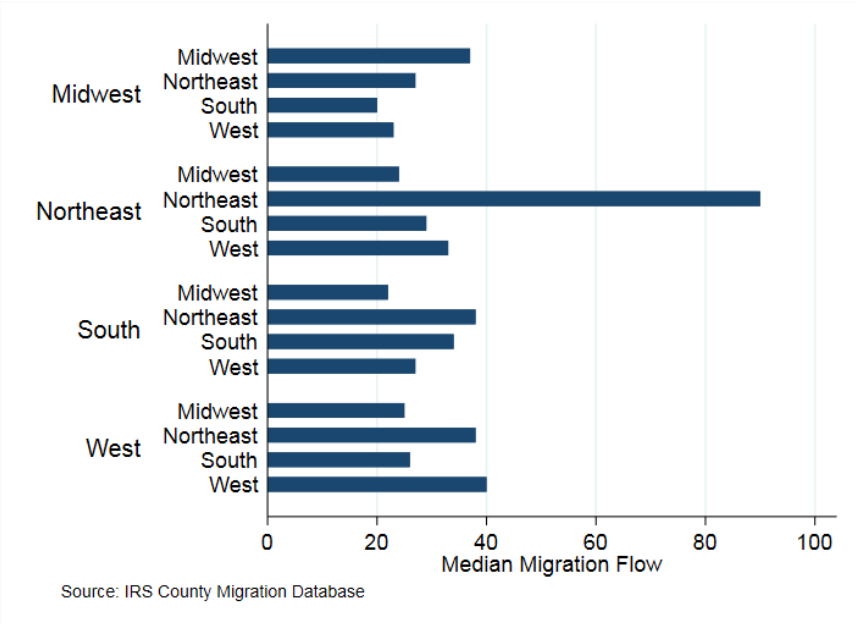
1.5 Figures and Tables

FIGURE 1.1: US imports from China in values and shares from 1991 to 2007



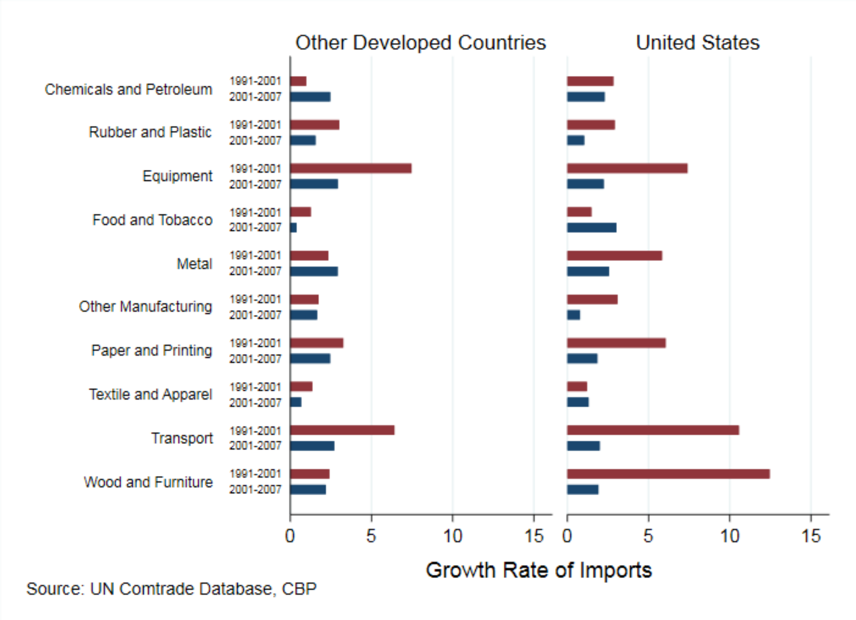
Notes: Import data comes from the UN Comtrade Database. China's share in US imports is calculated by dividing imports from China over imports from all countries. Import values are measured in 2007 dollars.

FIGURE 1.2: Migration between and within Census regions, 1992-2007



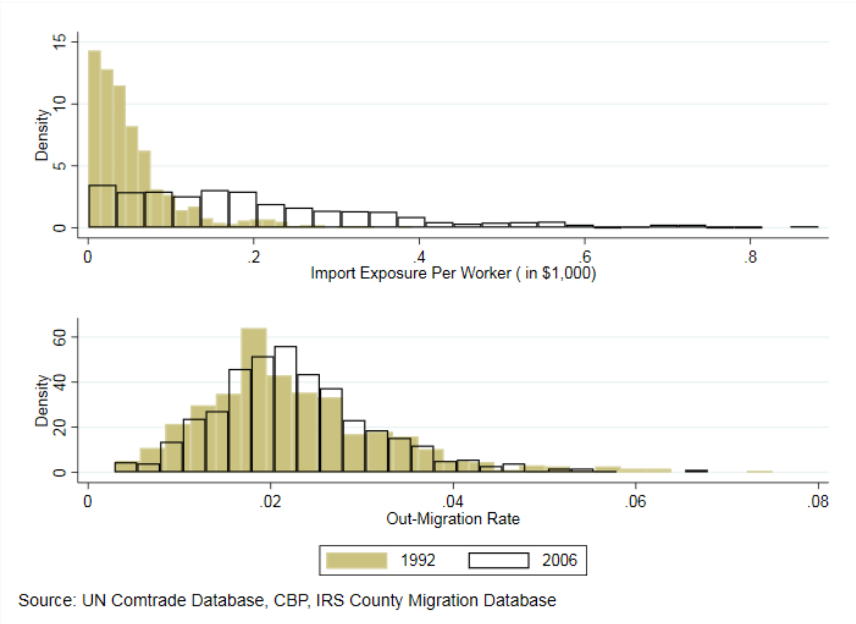
Notes: Data source is the county migration database from the Internal Revenue Service. This graph shows the median migration flows out of all flows within the same region, or between different regions from 1992 to 2007.

FIGURE 1.3: Growths in Imports from China for the US and other developed countries in 10 manufacturing industries, 1991-2007



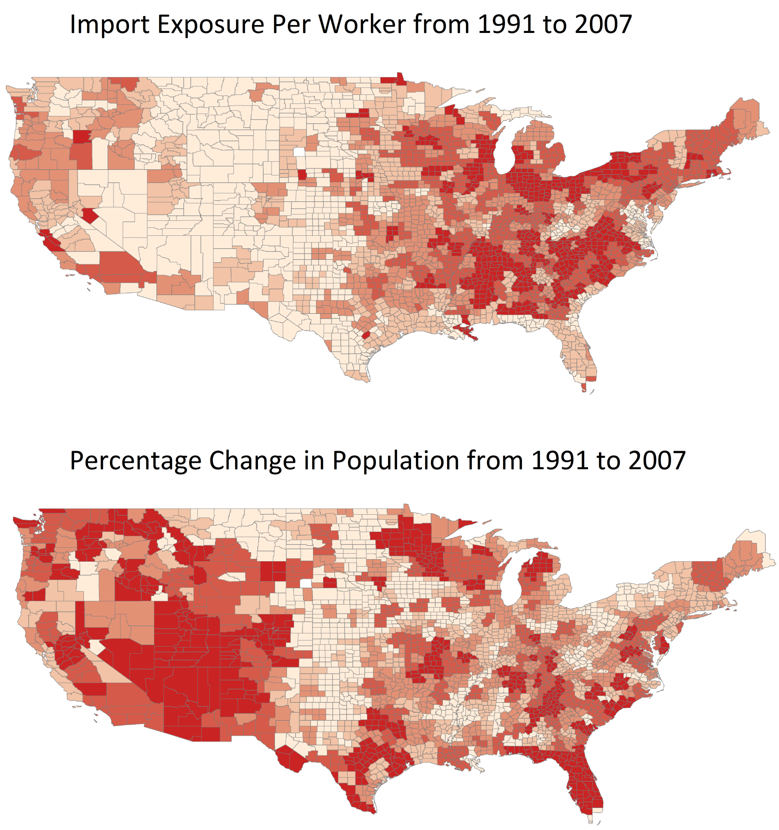
Notes: Import data comes from the UN Comtrade Database. The eight high-income countries are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland, for which disaggregated imports data since 1991 are available. Employment data comes from the CBP, which stands for County Business Patterns.

FIGURE 1.4: Distributions of import exposure per worker and out-migration rate across Commuting Zones in 1992 and 2006



Notes: Migration data comes from the IRS county migration database, and import data comes from the UN Comtrade Database. Employment data comes from the CBP, which stands for County Business Patterns. Import exposure per worker measures the intensity of import competition for each worker in a commuting zone. Out-Migration Rate is calculated by dividing the total migrants out of a commuting zone over its population.

FIGURE 1.5: Import Exposure Per Worker and Population Growth in US Commuting Zones from 1991 to 2007



Notes: Import exposure per worker measures the intensity of import competition for each worker in a commuting zone. Population growth is calculated using tax returns data from the IRS County Income Database. Different colors represent quantiles for each variable, with the darkest color representing the highest quantile and lightest representing the lower quantile.

TABLE 1.1: Summary Statistics of Migrant Flows and CZ Population

Year	N	Min	P10	Median	Mean	P90	Max
Panel A: Migrant Flows between Commuting Zones							
1992	22,278	10	11	29	117	231	19,702
1993	21,989	10	11	30	119	235	19,274
1994	22,498	10	11	30	120	237	19,405
1995	22,151	10	11	30	119	242	19,264
1996	22,249	10	11	31	123	246	19,934
1997	22,322	10	11	31	125	252	20,348
1998	22,539	10	11	31	128	251	20,766
1999	22,526	10	11	31	131	258	22,055
2000	22,489	10	11	32	134	264	22,230
2001	22,273	10	11	31	132	260	22,897
2002	21,684	10	11	31	131	257	22,677
2003	21,502	10	11	32	137	270	23,099
2004	22,080	10	11	32	135	268	23,464
2005	22,705	10	11	32	141	275	23,040
2006	22,468	10	11	32	136	266	21,684
Panel B: Commuting Zone Population							
1992	741	470	4,317	35,615	133,066	273,900	5,332,002
1993	741	458	4,317	35,849	133,888	276,615	5,208,509
1994	741	462	4,517	35,974	135,482	281,373	5,189,619
1995	741	449	4,488	36,491	137,724	287,258	5,258,566
1996	741	457	4,432	37,288	140,610	291,477	5,422,485
1997	741	456	4,560	37,772	143,121	299,191	5,560,714
1998	741	449	4,450	38,142	145,834	298,682	5,737,634
1999	741	466	4,440	38,571	148,203	307,082	5,874,130
2000	741	454	4,420	39,633	159,709	335,507	5,979,819
2001	741	442	4,425	38,970	151,082	319,593	6,100,189
2002	741	443	4,357	39,583	151,351	322,840	6,152,876
2003	741	432	4,354	39,721	152,222	326,756	6,220,849
2004	741	442	4,327	40,402	153,924	329,289	6,290,227
2005	741	435	4,234	40,681	154,989	335,926	6,346,745
2006	741	417	4,358	42,127	159,390	348,381	6,483,234

Notes: Migrant flows come from the IRS migration data. Commuting Zone population is tabulated from the County Income data. The unit of analysis is household. Migration flows that are less then 10 households are censored to preserve confidentiality.

TABLE 1.2: Summary Statistics of Average Migrant Income and Average CZ Income

Year	N	Min	P10	Median	Mean	P90	Max
Panel A: Average Migrant Income							
1992	22,278	1,398.72	21,875.96	33,815.37	38,121.60	57,586.08	1,224,927.97
1993	21,989	682.32	21,699.12	33,490.73	37,661.10	57,428.91	1,487,675.87
1994	22,498	486.13	22,057.94	33,832.45	37,927.44	57,543.09	1,073,383.55
1995	22,151	935.47	23,494.41	35,797.23	40,224.04	60,450.90	1,063,308.71
1996	22,249	769.42	23,432.28	36,266.85	41,462.27	63,208.11	1,967,145.84
1997	22,322	1,350.65	24,500.46	37,968.33	44,458.15	67,361.92	9,194,977.75
1998	22,539	1,058.54	25,474.93	40,052.20	46,973.80	71,432.15	4,913,597.13
1999	22,526	3,328.81	25,963.28	40,996.44	49,081.30	75,103.77	4,505,609.67
2000	22,489	3,000.06	26,245.54	41,726.31	50,090.44	78,086.14	2,984,606.22
2001	22,273	2,507.79	25,739.97	40,191.43	47,046.69	70,913.79	5,913,768.65
2002	21,684	970.43	25,354.57	38,998.79	44,003.14	66,910.82	1,299,731.00
2003	21,502	6,851.51	24,810.75	38,928.48	44,550.12	66,757.89	4,583,656.88
2004	22,080	4,570.64	25,400.21	39,608.11	45,390.11	68,714.70	1,650,892.39
2005	22,705	1,734.21	25,674.45	40,023.66	46,259.86	69,814.73	1,913,953.89
2006	22,468	4,222.57	26,074.38	40,428.02	47,041.94	70,206.77	3,290,861.30
Panel B: Average Commuting Zone Income							
1992	741	14,489.29	30,797.90	38,193.41	38,900.41	47,962.39	69,900.02
1993	741	20,229.47	31,132.13	38,134.17	38,891.70	47,636.42	70,809.90
1994	741	20,700.45	31,023.70	38,552.65	39,289.52	48,374.30	74,348.74
1995	741	20,859.70	31,486.14	39,482.74	40,103.54	49,850.78	75,479.31
1996	741	21,059.18	31,755.27	40,270.76	41,014.59	50,873.43	98,939.10
1997	741	10,213.43	32,688.28	41,287.39	41,911.72	52,518.68	92,634.65
1998	741	21,868.27	34,035.54	43,471.28	44,773.58	55,999.20	136,139.41
1999	741	23,531.86	35,583.43	44,830.44	46,250.88	58,152.64	122,490.65
2000	741	24,499.77	36,404.98	45,494.41	47,225.23	59,183.25	122,597.23
2001	741	25,149.94	35,751.57	44,187.44	45,604.21	56,987.75	109,549.55
2002	741	22,250.42	34,798.47	43,311.39	44,490.91	55,982.75	95,086.67
2003	741	25,346.21	35,088.16	43,508.56	44,780.66	55,612.84	89,555.06
2004	741	26,032.94	36,064.37	44,663.16	46,106.19	57,394.91	103,161.43
2005	741	27,062.27	37,101.87	45,705.40	47,193.79	58,754.63	98,340.99
2006	741	26,037.32	37,544.94	46,498.36	48,025.00	60,192.26	116,022.65

Notes: Migrant income data is extracted from the IRS County Migration Database. Commuting Zone average income is tabulated from county income data. All incomes are household gross adjusted incomes in 2009 dollars.

TABLE 1.3: Gravity estimates of US internal migration flows

	OLS				PPML		Tobit	
	1	2	3	4	1	2	1	2
Dependent Variable: Log Migrant Flows								
Log Origin Income	0.151***	0.435***	0.172***	-0.0202	0.152***	0.188***	0.186***	0.285***
	(0.0177)	(0.00141)	(0.0176)	(0.0136)	(0.0534)	(0.0618)	(0.0203)	(0.0179)
Log Dest Income	0.344***	0.449***	0.333***	0.159***	0.383***	0.465***	0.312***	0.529***
	(0.0179)	(0.00142)	(0.0178)	(0.0135)	(0.0389)	(0.0481)	(0.0124)	(0.0176)
Log Distance	-1.355***	-1.003***	-1.210***	-1.355***	-1.303***	-1.732***	-1.257***	-2.186***
	(0.00306)	(0.00289)	(0.00239)	(0.00306)	(0.00491)	(0.00544)	(0.00352)	(0.00301)
Origin FE	Y	N	Y	Y	Y	Y	Y	Y
Dest FE	Y	N	Y	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	N	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y	Y
N	321,773	321,773	333,753	321,773	321,773	8,092,350	321,773	8,092,350
Adjusted R sq.	0.600	0.445	0.582	0.599	N/A	N/A	N/A	N/A

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. The first four columns display results from OLS estimation of the gravity equation, which differ by fixed effects used. Results in column 1 for both PPML and Tobit estimations are based on positive migrant flows only, which are comparable to OLS in terms of the sample size. Results in column 2 for PPML and Tobit estimations are based on flows that are either positive or zero, and the sample covers all possible combinations of two CZs. Robust standard errors are reported in parentheses.

TABLE 1.4: Gravity estimates of US migrant incomes

	OLS				PPML		Tobit	
	1	2	3	4	1	2	1	2
Dependent Variable: Log Migrant Incomes								
Log Origin Income	0.355***	0.539***	0.382***	0.317***	0.412***	0.451***	0.379***	0.323***
	(0.0199)	(0.00154)	(0.0197)	(0.0156)	(0.0500)	(0.0570)	(0.0325)	(0.00438)
Log Dest Income	0.522***	0.513***	0.506***	0.480***	0.576***	0.656***	0.517***	0.332***
	(0.0205)	(0.00154)	(0.0203)	(0.0155)	(0.0405)	(0.0477)	(0.0437)	(0.00430)
Log Distance	-1.340***	-1.007***	-1.207***	-1.339***	-1.192***	-1.520***	-1.281***	-1.912***
	(0.00332)	(0.00303)	(0.00258)	(0.00332)	(0.00532)	(0.00619)	(0.00257)	(0.00995)
Origin FE	Y	N	Y	Y	Y	Y	Y	Y
Dest FE	Y	N	Y	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	N	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y	Y
N	321,773	321,773	333,753	321,773	321,773	8,092,350	321,773	8,092,350
Adjusted R sq.	0.596	0.469	0.581	0.595	N/A	N/A	N/A	N/A

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. The first four columns display results from OLS estimation of the gravity equation, which differ by fixed effects used. Results in column 1 for both PPML and Tobit estimations are based on positive migrant flows only, which are comparable to OLS in terms of the sample size. Results in column 2 for PPML and Tobit estimations are based on flows that are either positive or zero, and the sample covers all possible combinations of two CZs. Robust standard errors are reported in parentheses.

TABLE 1.5: Determinants of gross migration flows

	1	2	3	4	5	6	7
Dependent Variable: Log Gross Migration Flows							
Log Origin Population	0.168***	0.501***	0.170***	0.157***	0.103***	0.175***	0.171***
	(0.0212)	(0.00210)	(0.0207)	(0.0210)	(0.0283)	(0.0208)	(0.0212)
Log Origin Avg Income	-0.0614	-0.00962	0.0747**	-0.322***	-0.00907	0.198***	-0.187***
	(0.0512)	(0.0157)	(0.0378)	(0.0404)	(0.0688)	(0.0449)	(0.0506)
Log Dest Population	0.318***	0.507***	0.262***	0.313***	0.205***	0.318***	0.301***
	(0.0205)	(0.00209)	(0.0163)	(0.0204)	(0.0276)	(0.0203)	(0.0205)
Log Dest Avg Income	0.548***	0.0394**	0.601***	0.296***	0.177***	0.353***	0.613***
	(0.0510)	(0.0154)	(0.0424)	(0.0397)	(0.0667)	(0.0454)	(0.0504)
Log Distance	-1.362***	-1.026***	-1.215***	-1.361***		-1.356***	-1.361***
	(0.00307)	(0.00290)	(0.00240)	(0.00307)		(0.00306)	(0.00307)
Log Origin HPI	0.250***	0.429***	0.234***	0.225***	0.126***		0.233***
	(0.0188)	(0.0169)	(0.0181)	(0.0167)	(0.0248)		(0.0187)
Log Dest HPI	-0.173***	0.327***	-0.173***	-0.202***	-0.0820***		-0.158***
	(0.0189)	(0.0172)	(0.0187)	(0.0167)	(0.0247)		(0.0189)
Origin UR	2.578***	0.180*	2.460***	2.021***	1.175***	2.289***	
	(0.170)	(0.101)	(0.165)	(0.154)	(0.221)	(0.164)	
Dest UR	-1.593***	-2.407***	-1.564***	-2.194***	-0.905***	-1.423***	
	(0.174)	(0.106)	(0.169)	(0.156)	(0.224)	(0.168)	
Origin FE	Y	N	Y	Y	Y	Y	Y
Dest FE	Y	N	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	N	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y
N	315,655	315,655	327,068	315,655	315,655	321,773	315,655
Adjusted R sq.	0.601	0.453	0.583	0.601	0.290	0.600	0.601

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics. Robust standard errors are reported in parentheses.

TABLE 1.6: Determinants of net migration flows – *including* single direction flows

	1	2	3	4	5	6	7
Dependent Variable: Net Migration Flows							
Log Origin Population	-0.269***	-0.295***	-0.422***	-0.261***	-0.238***	-0.257***	-0.270***
	(0.0384)	(0.00336)	(0.0380)	(0.0381)	(0.0399)	(0.0384)	(0.0389)
Log Origin Avg Income	-2.274***	0.0237	-1.437***	-2.199***	-2.360***	-1.095***	-2.698***
	(0.0958)	(0.0274)	(0.0831)	(0.0750)	(0.0986)	(0.0854)	(0.0949)
Log Dest Population	-0.0467	-0.277***	0.234***	-0.0383	0.0193	-0.0742**	-0.121***
	(0.0374)	(0.00345)	(0.0296)	(0.0372)	(0.0390)	(0.0375)	(0.0374)
Log Dest Avg Income	2.054***	0.576***	1.395***	2.140***	2.317***	0.975***	2.391***
	(0.0907)	(0.0265)	(0.0752)	(0.0722)	(0.0965)	(0.0825)	(0.0899)
Log Distance	-0.991	1.251	-1.854	-0.988		-0.635	-0.688
	(2.386)	(1.771)	(2.367)	(2.380)		(2.360)	(2.384)
Log Origin HPI	1.043***	0.660***	0.927***	1.066***	1.131***		0.992***
	(0.0340)	(0.0299)	(0.0330)	(0.0307)	(0.0351)		(0.0340)
Log Dest HPI	-1.016***	-0.992***	-0.946***	-0.991***	-1.085***		-0.948***
	(0.0332)	(0.0293)	(0.0323)	(0.0297)	(0.0349)		(0.0332)
Origin UR	8.631***	8.171***	9.069***	8.613***	9.574***	7.680***	
	(0.309)	(0.203)	(0.306)	(0.280)	(0.326)	(0.301)	
Dest UR	-7.626***	-2.585***	-8.108***	-7.651***	-8.116***	-6.860***	
	(0.317)	(0.191)	(0.306)	(0.282)	(0.336)	(0.311)	
Origin FE	Y	N	Y	Y	Y	Y	Y
Dest FE	Y	N	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	N	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y
N	188,268	188,268	195,070	188,268	188,268	192,211	188,268
Adjusted R sq.	0.378	0.218	0.365	0.378	0.309	0.372	0.374

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. The dependent variable used for estimations in this table includes net migration flows in which one direction is reported as zero flow. Net flows are calculated as the differences between log migration flows in both directions. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics. Robust standard errors are reported in parentheses.

TABLE 1.7: Determinants of net migration flows – *excluding* single direction flows

	1	2	3	4	5	6	7
Dependent Variable: Net Migration Flows							
Log Origin Population	-0.0835***	-0.00417**	-0.163***	-0.0866***	-0.0836***	-0.0844***	-0.0590**
	(0.0220)	(0.00199)	(0.0218)	(0.0218)	(0.0220)	(0.0223)	(0.0222)
Log Origin Avg Income	-1.247***	-0.143***	-0.759***	-1.301***	-1.247***	-0.510***	-1.483***
	(0.0540)	(0.0154)	(0.0497)	(0.0436)	(0.0540)	(0.0485)	(0.0540)
Log Dest Population	0.109***	0.000212	0.239***	0.106***	0.109***	0.102***	0.0575***
	(0.0218)	(0.00196)	(0.0170)	(0.0217)	(0.0218)	(0.0222)	(0.0220)
Log Dest Avg Income	1.392***	0.186***	1.009***	1.336***	1.392***	0.606***	1.636***
	(0.0529)	(0.0153)	(0.0433)	(0.0432)	(0.0529)	(0.0477)	(0.0528)
Log Distance	-2.130	-1.398	-1.521	-2.126		-2.091	-2.162
	(3.384)	(2.543)	(3.372)	(3.380)		(3.370)	(3.387)
Log Origin HPI	0.658***	0.455***	0.591***	0.667***	0.658***		0.611***
	(0.0193)	(0.0177)	(0.0185)	(0.0177)	(0.0193)		(0.0194)
Log Dest HPI	-0.693***	-0.446***	-0.646***	-0.684***	-0.693***		-0.645***
	(0.0190)	(0.0176)	(0.0182)	(0.0175)	(0.0190)		(0.0191)
Origin UR	5.532***	3.769***	5.734***	5.587***	5.533***	4.885***	
	(0.179)	(0.120)	(0.172)	(0.161)	(0.179)	(0.173)	
Dest UR	-5.604***	-3.576***	-5.817***	-5.546***	-5.604***	-4.901***	
	(0.180)	(0.121)	(0.171)	(0.163)	(0.180)	(0.175)	
Origin FE	Y	N	Y	Y	Y	Y	Y
Dest FE	Y	N	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	N	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y
N	127,387	127,387	131,998	127,387	127,387	129,562	127,387
Adjusted R sq.	0.387	0.185	0.379	0.387	0.387	0.375	0.377

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. The dependent variable used for estimations in this table includes net migration flows where flows in both directions positive. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics. Robust standard errors are reported in parentheses.

TABLE 1.8: Determinants of Migrants' Average Income

	1	2	3	4	5	6	7
Dependent Variable: Average Migrant Income							
Log Origin Population	0.0509***	0.0367***	0.0885***	0.0496***	0.0504***	0.0526***	
	(0.00887)	(0.000819)	(0.00866)	(0.00881)	(0.00877)	(0.00888)	
Log Origin Avg Income	0.699***	0.757***	0.430***	0.483***	0.744***	0.723***	
	(0.0232)	(0.00663)	(0.0190)	(0.0186)	(0.0203)	(0.0229)	
Log Dest Population	0.0318***	-0.0106***	0.113***	0.0314***	0.0334***	0.0290***	
	(0.00912)	(0.000895)	(0.00731)	(0.00908)	(0.00901)	(0.00909)	
Log Dest Avg Income	0.712***	0.765***	0.522***	0.498***	0.655***	0.739***	
	(0.0236)	(0.00704)	(0.0190)	(0.0184)	(0.0209)	(0.0232)	
Log Distance	0.0162***	0.00659***	0.00317***	0.0165***	0.0158***	0.0161***	0.0166***
	(0.00105)	(0.000922)	(0.000839)	(0.00105)	(0.00104)	(0.00105)	(0.00105)
Log Origin HPI	0.0457***	-0.0334***	0.0850***	0.0109		0.0478***	
	(0.00795)	(0.00648)	(0.00757)	(0.00726)		(0.00793)	
Log Dest HPI	-0.0543***	-0.0321***	-0.0422***	-0.0924***		-0.0516***	
	(0.00839)	(0.00698)	(0.00802)	(0.00770)		(0.00838)	
Origin UR	-0.373***	0.176***	-0.520***	-0.651***	-0.381***		
	(0.0716)	(0.0416)	(0.0692)	(0.0646)	(0.0696)		
Dest UR	-0.486***	-0.0239	-0.617***	-0.779***	-0.410***		
	(0.0730)	(0.0427)	(0.0701)	(0.0655)	(0.0709)		
Origin FE	Y	N	Y	Y	Y	Y	Y
Dest FE	Y	N	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	N	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y
N	315,655	315,655	327,068	315,655	321,773	315,655	321,773
Adjusted R sq.	0.458	0.370	0.450	0.457	0.465	0.458	0.459

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics. Robust standard errors are reported in parentheses.

TABLE 1.9: Average Migrant Income and the Migrant Share

	1	2	3	4	5
Dependent Variable: Average Migrant Income					
Migrant Share	-1.667***	-1.185***	-1.722***	-1.615***	-0.699***
	(0.261)	(0.348)	(0.262)	(0.262)	(0.302)
Log Distance		0.0179***			0.0170***
		(0.00130)			(0.00117)
Log Origin Population			0.0569***	0.0508***	0.0512***
			(0.00878)	(0.00888)	(0.00888)
Log Origin Avg Income			0.766***	0.699***	0.699***
			(0.0231)	(0.0232)	(0.0232)
Log Origin HPI			0.0402***	0.0471***	0.0456***
			(0.00776)	(0.00795)	(0.00795)
Origin UR			-0.453***	-0.355***	-0.374***
			(0.0703)	(0.0715)	(0.0716)
Log Dest Population				0.0331***	0.0318***
				(0.00913)	(0.00912)
Log Dest Avg Income				0.716***	0.712***
				(0.0236)	(0.0236)
Log Dest HPI				-0.0555***	-0.0542***
				(0.00839)	(0.00839)
Dest UR				-0.494***	-0.486***
				(0.0730)	(0.0730)
Origin FE	Y	Y	Y	Y	Y
Dest FE	Y	Y	Y	Y	Y
Region Pair FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
N	321,773	321,773	318,160	315,655	315,655
Adjusted R sq.	0.458	0.348	0.458	0.458	0.458

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics. Robust standard errors are reported in parentheses.

Chapter 2

Importing Migration? The Effects of Import Competition on Internal Migration Patterns in the United States

2.1 Introduction

This paper studies how rising import competition from China affects migration flows within the US. First, I present a theoretical framework that generates theoretical predictions on the marginal effects of import competition on migration flows. I then employ two-stage least squares strategy to identify the causal effects of rising import competition on migration flows, out-migration rates and local populations. The identification makes use of the fact that concurrent growths in imports of Chinese goods for non-US developed countries are driven by higher Chinese productivities and lower trade costs.

I extend the specific factors model in Kovak (2013) by adding endogenous sorting of labor and allowing labor supply to be elastic across regional economies. In this model, workers with heterogeneous productivity levels select into locations that offer higher wages and amenity values, and are in proximity to their previous locations. As in Bryan and Morten (2017), worker productivities follow the Frechet distribution, and migration flows can be written as a log-linear combination of wages and distances between local economies.[1] Average migrant income decreases with migrant share and increases with distance as a result of selection. These relationships provide the theoretical underpinnings of the gravity estimations.

In this model, intensifying import competition exerts downward pressure on industry prices, resulting in wage reductions and population declines in local economies, the extent of which depends on a variety of factors.[2] If an industry has a higher presence in a local economy, as measured by the local industry employment share, the same change in industry price will have a disproportionate impact on wage and population in that area. Meanwhile, the effects of industry price changes on wage and population are bigger for industries with more elastic labor demand, since these industries need to absorb/unleash more labor to the local economy to restore equilibrium when price rises/drops. On the labor supply side, when worker productivities are more dispersed or less correlated across destinations, industry price changes lead to bigger wage changes and smaller population changes.[3] Additionally, if worker utility declines faster with congestion, local labor supply will be less elastic and the impact of industry shocks will fall more on wages and less on populations.

To identify the causal effects of rising import competition from China on internal migration flows in the United States, I use the instrumental variable

[1] The assumption on worker productivity follows from Hsieh et al. (2016), which in turn borrows from Eaton and Kortum (2002).

[2] Each local economy is assumed to have full employment, so there is no distinction between employment and population in the model.

[3] In other words, when the role of comparative advantage is more important, less workers relocate in response to industry shocks.

strategy inspired by Autor, Dorn, and Hanson (2013). The main estimating equations relate population, out-migration rate and bilateral migration flows to measures of import exposure for both the origin and the destination CZ.[4] The idea of the import exposure variable is to allocate changes in national industry imports to the CZ level, using CZ employment shares within each industry. To isolate the effects of industry demand shocks in the US that simultaneously affect local labor demand and industry imports, I construct an instrument for the import exposure variable using imports of other developed countries from China. The instrumental variable captures exogenous variations in Chinese productivities and trade costs that drive import growths across developed markets, instead of demand shocks specific to the US.

Results show that a one standard deviation increase in imports per worker in a given location lowers the local population by 1.34 percent, while raising the out-migration rate by 0.28 percentage points in a CZ. A one standard deviation increase in imports per worker in the *origin* CZ raises net migrant outflows by 0.11 log points, while the same increase in the *destination* CZ lowers net inflows by 0.08 log points. These estimates point to the fact that import competition affects net migration flows, and the effect is stronger in the *origin* than in the *destination*. Such disparity may stem from information asymmetry of potential migrants who are more familiar with their current residences than future ones.

This paper contributes to two strands of literature. First, numerous studies have analyzed the labor market outcomes of trade and offshoring across industries and occupations, but relatively few have focused on the geographic aspect of labor adjustment, especially in the context of rising imports from

[4]For outcomes at the CZ level (population and out-migration rate), import exposure in the CZ is used as the explanatory variable; for outcomes at the CZ pair level (gross and net migration flows), import exposure for the origin and destination CZs are used. Import competition and import exposure are used interchangeably in this paper.

China.[5] For example, Pierce and Schott (2012) find that US industries subject to greater potential tariff hikes experienced greater employment losses after 2000. In a study that compares occupation with industry responses to trade and offshoring, Ebenstein et al. (2014) find slower wage growths in US occupations that were more exposed to import competition from low-wage countries, and that occupation tenure is more important than industry tenure in determining wage outcomes. Using administrative employment data from Brazil, Dix-Caneiro and Kovak (2017) find sustained employment and wage effects in Brazilian regions facing tariff reductions, suggesting slow adjustment of labor across regions. Sluggish mobility response to labor demand shocks is also found for the US (Topel (1986); Blanchard and Katz (1992); Glaeser and Gyourko (2005)), though there lacks direct evidence on the rate of adjustment in response to import competition. This paper aims to fill this gap by providing estimates on the elasticity of migration with respect to rising import competition from China.

A closely related paper by Autor, Dorn, and Hanson (2013) find that over 1990-2007, intensifying import competition from China lowered wages and raised unemployment rates across local labor markets in the US. However, they find no significant effect on the local population using data from the Census and the ACS. This study uses tax returns information from the IRS that is less susceptible to sampling and measurement errors, and finds significant effects of import exposure on CZ population, measured by the number of tax returns or exemptions filed. Moreover, the richness and the disaggregated nature of the IRS migration data allow estimation of the effects on bilateral flows between local labor markets, which also turn out to be significant.

[5] Hakobyan and McLaren (2016) analyze the effects of North American Free Trade Agreement on industries and localities in the US, and find significant wage effects for blue-collar workers in the most affected industries and localities. Greenland, Lopresti, and McHenry (2017) look at how potential tariff hikes affected US internal migration and find delayed response to import competition.

Second, this paper adds to the literature on US migration by offering new evidence on the determinants of migration patterns. Existing research on US internal migration patterns have documented a secular decline in interregional migration rates, and sought to explain such phenomenon. For example, by comparing inter-state migration rates across three different data sources – the Census, the CPS and the IRS, Molloy, Smith, and Wozniak (2011) find average migration rates to have steadily declined since the 1980's. Their estimates show a large disparity between estimates using the March CPS versus estimates using the Census or the IRS migration data, which they attribute to a much smaller sample size of the March CPS. Kaplan and Schulhofer-Wohl (2017) argue that declining migration rates result from less geographic dispersion of wages within occupations and better access to information about potential destinations.

Despite evidence on declining migration rates nationally, migration patterns on the sub-national level are more varied and some areas may experience higher migration due to labor demand shocks. Partridge et al. (2012) find that industry employment growth, a proxy for demand shocks, affect interregional migration more in the pre-2000 period than in the post-2000 period. However, the geographic units of migration flows they construct from the Census and the CPS are states and metropolitan areas; therefore migration within states and outside of metro areas are not included in the analyses. Additionally, the industry employment growth variable may be endogenous to factors that simultaneously influence industry employment and migration, as some industries tend to cluster in a particular region. On the other hand, this paper uses migration data at the most disaggregated level of geographic classification, for every pair of locations, and exploits exogenous variations in Chinese productivities and trade costs in identifying the causal effect of rising import competition on migration.

The remainder of this paper is organized as follows. Section 2.1 lays out the theoretical framework. Section 2.2 discusses the empirical strategy. Section 2.3 presents empirical results and Section 2.4 shows additional results for robustness. Section 2.5 concludes and discusses areas for future research.

2.2 Theoretical Framework

In this section, I present a spatial equilibrium model that accounts for the internal migration patterns, and guides the empirical strategy in identifying the effects of import competition on migration and population. This framework reconciles the empirical finding on higher mobility induced by rising imports, by extending the specific factors model considered in Kovak (2013) and allowing regional wages to be determined by the endogenous supply of workers in addition to industry prices. In this model, workers with heterogeneous productivities select into locations that offer different wage and amenity levels. The effect of industry price changes on local wage and population not only depends on local industry composition and labor demand elasticity, but parameters that govern the elasticity of labor supply. The extension on endogenous labor supply borrows from the worker selection model in Bryan and Morten (2017), which builds upon Hsieh et al. (2016) and Eaton and Kortum (2002).

Consider an economy consisting of N regions. There are J industries within each region, and the production of each industry utilizes two inputs, a mobile input – Human Capital (H) and an immobile input (T).[6] Human capital freely moves across industries within the same region, since manual skills for manufacturing industries considered in this paper are transferable to another industry. However, it is costly for workers (human capital) to move to

[6] Human capital is interchangeable with effective labor units supplied by the worker. Workers with higher productivity supply more effective labor units.

another region, and the cost increases with the distance traveled. The immobile factor is specific to the industry, and determines the labor productivity in that industry. Examples of the immobile factor include natural resources, fixed capital and existing industry agglomeration. Thus the model accounts for labor market outcomes in the short run. Assume production exhibits constant returns to scale. Production technologies are different across industries, but are the same for each region within an industry as all regions have the same access to technologies.

2.2.1 Production

Let $d \in \{1, ..., N\}$ index each region, and $j \in \{1, ..., J\}$ index each industry. Output for industry j in area d is denoted as Y_{jd}. Workers supply human capital H_d to industries in area d. The immobile input T_{jd} is fixed for an industry–area. Unit input requirements for H_{jd} and T_{jd} are a_{Hj} and a_{Tj} respectively. The factor market clearing conditions are

$$a_{Tj} Y_{jd} = T_{jd} \tag{2.1}$$

$$\sum_j a_{Hj} Y_{jd} = H_d \tag{2.2}$$

Total differentiation of these equations yields

$$\hat{Y}_{jd} = -\hat{a}_{Tj} \tag{2.3}$$

and

$$\sum_j \varphi_j (\hat{a}_{Hj} - \hat{a}_{Tj}) = \hat{H}_d \tag{2.4}$$

where $\varphi_{jd} = \frac{H_{jd}}{H_j}$ is the fraction of human capital utilized in area d industry j over total human capital in industry j. Under perfect competition, output price is equal to total factor payments, thus

$$a_{Hj} w_d + a_{Tj} R_{jd} = P_j \quad \forall j \tag{2.5}$$

where w_d and R_{jd} are prices of human capital and immobile input, and P_j is the price of industry j output. Denote β_j as the cost share of the immobile input. Industry price changes can be written as

$$(1 - \beta_j)\hat{w}_d + \beta_j \hat{R}_{jd} = \hat{P}_j \quad \forall j \tag{2.6}$$

which follows from the relationship between input requirements implied by the envelope theorem:

$$(1 - \beta_j)\hat{a_{Hj}} + \beta_j \hat{a_{Tj}} = 0 \quad \forall j \tag{2.7}$$

Let σ_j denote the elasticity of substitution between H and T. The definition of σ_j implies

$$\hat{a_{Tj}} - \hat{a_{Hj}} = \sigma_j(\hat{w}_d - \hat{R}_{jd}) \tag{2.8}$$

Plugging equation (2.8) into (2.4), I obtain

$$\sum_j \varphi_{jd} \sigma_j (\hat{R}_{jd} - \hat{w}_d) = \hat{H}_d \tag{2.9}$$

The solution for $\hat{w}_d$ using equations (2.6) and (2.9) is

$$\hat{w}_d = \frac{\hat{H}_d}{\sum_{j'} \varphi_{j'd} \frac{\sigma_{j'}}{\beta_{j'}}} + \sum_j \frac{\varphi_{jd} \frac{\sigma_j}{\beta_j}}{\sum_{j'} \varphi_{j'd} \frac{\sigma_{j'}}{\beta_{j'}}} \hat{P}_j \tag{2.10}$$

Suppose labor is supplied inelastically to each area, equation (2.10) shows that the effect of industry prices changes on local wages depends on φ_{jd} – the local industry employment share, and $\frac{\sigma_j}{\beta_j}$ – the demand elasticity for human capital.[7] If an area has a large presence of a particular industry, then local wages will be more affected by price changes of this industry. If an industry has more elastic demand for human capital, then a price decrease means that industry has to unleash more workers into the local area to restore equilibrium, resulting in bigger wage reductions. However, the model considered in this paper allows for elastic labor supply, so it is unclear whether to what extent these insights are preserved or additional parameters need to be taken into account. The following subsection describes migration and labor supply.

2.2.2 Migration and Labor Supply

Let o denote the origin and d the destination. Each worker i in their origin receives a skill draw from a multivariate Frechet Distribution for a potential destination. The CDF of this Frechet distribution is given as

$$F(s_1, ..., s_N) = exp-[\sum_{d=1}^{N} s_d^{-\frac{\tilde{\theta}}{1-\rho}}]^{1-\rho}$$

In this expression, $\tilde{\theta}$ indicates the extent of skill dispersion across destinations. A higher $\tilde{\theta}$ means there's less dispersion across skills in different destinations. If comparative advantage plays an important role in determining migration, $\tilde{\theta}$ is expected to be small. ρ measures the correlation between skills in different destinations. ρ is higher when workers are more likely to be productive in d' if they are productive in d, therefore it captures how much "general talent" matters for the skill draw. For simplicity, let $\theta = \frac{\tilde{\theta}}{1-\rho}$.

In addition to skills, education quality in the origin is another factor that determines the amount of human capital workers supply.[8] Let q_o denote the

[7] $\frac{\sigma_j}{\beta_j}$ represents the industry demand elasticity for human capital when T_jd is fixed.

[8] The amount of human capital a worker supplies is equal to her productivity.

quality of human capital formation, then the total human capital (or effective labor units) for worker i moving from o to d is

$$h_{ido} = s_{id}q_o$$

and her "effective" wage is

$$w_{ido} = w_d h_{ido} = w_d s_{id} q_o$$

Let τ_{do} denote the distance between o and d. Worker utility is

$$U_{do} = \alpha_d(1 - \tau_{do})w_d s_{id} q_o$$

where $\alpha_d = \bar{\alpha_d} L_d^{\lambda}$ is the endogenously determined amenity value for the destination. $\bar{\alpha_d}$ is the intrinsic amenity value for an area, which reflects how attractive a location is for its residents. For example, climate can determine the intrinsic attractiveness of an area. Overall amenity is also affected by the number of workers/residents in the area. λ measures the extent of congestion in terms of overall amenity, and is assumed to be negative. If more people move to a destination, utility for every worker in that destination will decline, the rate of which is governed by λ. τ_{do} represents the cost to utility when a worker migrates to another area, which is proxied by distance. If the worker stays in the origin, then τ_{do} is equal to 0. Migration cost is assumed to be symmetric, i.e. $\tau_{do} = \tau_{od}$. The expression for utility implies that workers with better skill draws are more likely to move to locations farther away, holding all others constant. The share of migrants from o to d can be written as

$$\pi_{do} = \frac{\tilde{w}_{do}^{\theta}}{\sum_{l=1}^{N} \tilde{w}_{lo}^{\theta}} \tag{2.11}$$

where $\tilde{w}_{jo} = w_d \alpha_d (1 - \tau_{do})$. Migrant share increases with wage and amenity in the destination, while decreases with the distance between the origin and the destination. Therefore, net migration flows from the origin to the destination are expected to increase with wages and amenity values in the destination, while decrease with those in the origin. Distance does not have any effect on net flows since the effects in one direction and the opposite direction cancel out. However, the effect of distance on gross migration flows are likely to be negative, while wages and amenity have similar effects as on net flows. These predictions are consistent with the empirical observations. The expected value of skills from o to d is

$$E[s_{do} \mid d] = (\frac{1}{\pi_{do}})^{\frac{1}{\theta}}\bar{\Gamma}$$

where $\Gamma = \Gamma(1 - \frac{1}{\theta(1-\rho)})$ is the Gamma function. Then the expected wage for migrants from o to d is

$$\bar{w}_{do} = w_d q_o E[s_{do} \mid d] = w_d q_o (\frac{1}{\pi_{do}})^{\frac{1}{\theta}}\bar{\Gamma}$$

According to this expression, average migrant wage is positively related to the destination wage and negatively related to the migration share. For a pair of locations that are far apart, workers with higher skill draws in the origin are more likely to move. The greater the distance, the lower the share of workers who move to the destination. Therefore, migrant productivity and average wage increase with the distance between o and d. This selection result is again confirmed in the data. Total human capital supplied at d is given by

$$H_d = \sum_o q_o \bar{L}_o \pi_{do} E[s_{do} \mid d] = \sum_o q_o \bar{L}_o \pi_{do} (\frac{1}{\pi_{do}})^{\frac{1}{\theta}}\bar{\Gamma}$$

Substituting π_{do} using equation (2.11), the expression for H_d becomes

$$H_d = \bar{\alpha_d}^{(\theta-1)} w_d^{(\theta-1)} L_d^{\lambda(\theta-1)} \sum_o \frac{q_o \bar{L_o} \Gamma(1-\tau_{do})^{(\theta-1)}}{\Theta^{(1-\frac{1}{\theta})}} \tag{2.12}$$

where $\Theta = \sum_{l=1}^{N} \tilde{w}_{lo}^{\theta}$. Log-linearizing this equation yields

$$\hat{H_d} = (\theta-1)\hat{w}_d + \lambda(\theta-1)\hat{L_d} \tag{2.13}$$

2.2.3 Equilibrium

In equilibrium, workers endowed with heterogeneous productivities select into locations that offer the highest returns on utility, taking into account wages, amenity, and migration costs. Industries in each location choose the amount of human capital to maximizes profits given the exogenous industry price and fixed amount of the specific factor, T_{jd}. The equilibrium is characterized by the following conditions:

1. Consumers maximize utility

$$\pi_{do} = \frac{\tilde{w}_{do}^{\theta}}{\sum_{j=1}^{N} \tilde{w}_{jo}^{\theta}} \tag{2.14}$$

2. Producers maximize profit

$$\hat{w}_d = \frac{\hat{H}_d}{\sum_{j'} \varphi_{j'd} \frac{\sigma_{j'}}{\beta_{j'}}} + \sum_j \frac{\varphi_{jd} \frac{\sigma_j}{\beta_j}}{\sum_{j'} \varphi_{j'd} \frac{\sigma_{j'}}{\beta_{j'}}} \hat{P}_j \tag{2.15}$$

where $\hat{H_d} = (\theta-1)\hat{w}_d + \lambda(\theta-1)\hat{L_d}$.

3. Labor markets clear

$$L_d = \sum_o \bar{L}_o \pi_{do} \tag{2.16}$$

Plugging equation (12) into (14), I obtain

$$L_d = \sum_o \bar{L}_o \pi_{do} = \bar{\alpha_d}^{\theta} w_d^{\theta} L_d^{\lambda\theta} \sum_o \frac{q_o \bar{L_o}(1-\tau_{do})^{\theta}}{\Theta}$$

Log-linearizing this equation gives

$$\hat{L_d} = \frac{1}{\frac{1}{\theta} - \lambda} \hat{w}_d \tag{2.17}$$

Equation (2.17) characterizes the positive relationship (since $\lambda < 0$) between changes in labor supply and wage level at the destination. Recall that θ increases with the correlation and decreases with the dispersion in skills across locations. A bigger θ means the same change in wage results in a bigger change in the labor supply, when skills are more correlated or less dispersed across destinations. This is due the fact that when worker productivities are more "homogeneous", a small increase in w_d makes d more attractive to many other workers who would have chosen somewhere else. In a case where the gaps between worker productivities are large, it requires bigger changes in w_d to motivate the same number of workers to move to d.

Additionally, the same change in wage is associated with a smaller change in labor supply if the congestion parameter λ is bigger in magnitude. This negative parameter determines how fast amenity in a location declines when more workers move in. If the congestion force is strong (i.e. $\lambda \ll 0$), less workers will have enough increase in utility that makes them move elsewhere, amounting to a smaller change in labor supply at any location.

To gauge the effect of industry price changes on population and migration, I rewrite equation (2.13), replacing $\hat{H_d}$ with the expression of $\hat{w}_d$ and $\hat{L_d}$

$$(1 + \frac{\theta - 1}{C_j})\hat{w}_d = -\frac{\lambda(\theta - 1)}{C_j}\hat{L_d} + \sum_j \xi_j \hat{P}_j \tag{2.18}$$

where $C_{jd} = \sum_{j'} \varphi_{j'd} \frac{\sigma_{j'}}{\beta_{j'}}$ and $\xi_{jd} = \frac{\varphi_{jd} \frac{\sigma_j}{\beta_j}}{\sum_{j'} \varphi_{j'd} \frac{\sigma_{j'}}{\beta_{j'}}}$.

Combining equation (2.17) with (2.18), I obtain the following results

$$\hat{L_d} = \hat{\pi}_{do} = \frac{\sum_j \xi_{jd} \hat{P}_j}{(1 - \frac{1}{C_{jd}})\frac{1}{\theta} - \lambda + \frac{1}{C_{jd}}} \tag{2.19}$$

$$\hat{w}_d = \frac{\sum_j \xi_{jd} \hat{P}_j}{1 + \frac{\theta - 1}{(1-\lambda\theta)C_{jd}}} \tag{2.20}$$

In equations (2.19) and (2.20), the marginal effects of price changes on the outcome variables are positive as long as $\theta > \frac{1-C_j}{1-C_j\lambda}$. In fact, the positive relationships hold if $\theta > 1$, which, according to Bryan and Morten (2017) is true as they estimated θ to be 28 for the US, and 13 for Indonesia. Estimates on the marginal effects from this paper also confirm the positive sign of the coefficient.

The magnitude of the price effects depends on the industry composition as well as parameters that govern the elasticity of labor demand and supply. A higher presence of a particular industry in a local labor market means the effects on population, migrant share and wage are stronger when the price of that industry falls. If labor demand is more elastic, the effects are also stronger. Furthermore, if worker skills are more correlated or less dispersed across locations, the same changes in industry prices will result in bigger changes in population and migrant share, and smaller changes in wage. On the other hand, if worker utility declines faster due to congestion, the impact of industry price shocks will fall more on wages and less on population and migrant share, as labor supply is less elastic. By incorporating endogenous labor supply into the specific factors model, this new framework generates theoretical predictions on how industry price changes affect population and migration, and offers fresh insights on the parameters that determine the magnitude of the effects.

2.3 Empirical Strategy

In this section, I describe the mapping from theoretic predictions to empirical estimations, and how the empirical strategy accounts for potential endogeneity concerns. According to equation (2.19), there is a linear relationship between industry price changes, and changes in population and migration rate. Industry price changes are proxied by increased imports from China, as trade liberalization leads to changes in both quantities and prices of imported products. However, using quantities as the measure of import competition instead of prices (or tariffs) may capture effects of trade liberalization resulted from the removal of non-tariff barriers, in addition to effects caused by lower tariffs.[9]

To map industry-level changes in imports to the local labor market, I multiply industry imports with employment shares of local areas within each industry. The summation of local import exposures by industry measures how local economies are affected by import shocks differentially according to the initial industry employment compositions and the extent of trade shocks.[10] The expression for the import exposure variable is as follows:

$$\Delta IME_{dt}^{US-China} = \sum_{j} \frac{E_{djt}}{E_{jt}} \frac{\Delta M_{jt}^{US-China}}{E_{dt}} \tag{2.21}$$

In this expression, d denotes local area, j denotes industry and t denotes the first year of the sample. $\frac{E_{djt}}{E_{jt}}$ is the fraction of employment in area d over total employment in industry j in the US. $\Delta M_{jt}^{US-China}$ represents the annual change in US imports from China in industry j between the starting year and the end year of the sample. This measure apportions a change in industry imports to each local area according to the share of industry employment

[9]For example, Harrigan and Burrows (2009) documented a 450% quantity increase in US imports of apparel and textiles from China after quota restrictions were eliminated in 2005.

[10]Import exposure and import competition are used interchangeably.

in that area, and then aggregates all local industry trade shocks. The variations in this measure come from two sources. First, given the same industry trade shock, the areas that represent more local industry employment over national industry employment at the starting year receive bigger shocks. Second, given the same industry employment shares within an area, the local industry trade shock is larger in industries that are exposed to bigger shocks on the national level.

Ideally, the import exposure variable captures exogenous variations in trade barriers and Chinese productivities that result in observed changes in imports. However, industry demand shocks may simultaneously affect imports as well as local economic conditions and migration patterns. This problem is more pronounced when industries are sometimes clustered in a particular region. To account for endogeneity, I employ an instrument that measures import exposure using imports of 11 other developed countries from China:[11]

$$\Delta IME_{dt}^{Other-China} = \sum_j \frac{E_{djt-1}}{E_{jt-1}} \frac{\Delta M_{jt}^{Other-China}}{E_{dt-1}} \tag{2.22}$$

In this expression, imports of other developed countries replace US imports as in the import exposure variable. To account for adjustments in industry employment in expectation of rising import competition, employment shares in this variable are from the previous decade. The validity of this instrument relies on the assumption that any import demand shock is specific to the importing country, but not the same across different developed markets. In other words, this instrument captures changes in China's manufacturing productivity and trade costs rather than the importing country's domestic economic conditions.

[11]The instrumental variable strategy follows from Autor, Dorn, and Hanson, 2013. However, this paper allows more time frequency as it considers changes in imports between two consecutive years, while Autor, Dorn, and Hanson, 2013 uses changes in imports over a decade.

The validity of the instrument is at stake, if common demand shocks across developed markets, instead of supply shocks from China are driving the variation in US migration and imports in other developed countries. For example, there was a housing boom across developed countries in the early 2000's due to favorable credit conditions, which may raise demand for products related to the housing market. To deal with this issue, I drop imports in industries that are related to housing (i.e. steel, cement and furniture) for the construction of the import exposure variable as well as the instrument. To account for the possibility that common technology shocks were raising demands for electronics across developed countries, I drop imports in industries such as electronic equipment. There may be additional concerns on the competition between Chinese exports and US exports to non-US developed markets that adversely affect US industries, which are addressed by using Iranian imports from China in the construction of the instrument. Due to economic sanctions including trade restrictions that the US imposed against Iran following the 1979 revolution, it is unlikely that US exports have been competing with Chinese exports in the Iranian market. Additionally, I use alternative measures for US imports, including net imports, imports net of imported intermediates and imports from other low wage countries to compare estimates. Discussions of results using alternative instruments and import measures are included in Section 2.4.

2.4 Estimation Results

I present estimates on population and migration from two-stage least squares estimation, using an instrument that captures exogenous variations in lower tariffs and higher Chinese productivities. I begin with results on the population, followed by results on the out-migration rate. To explore the effects of import competition on bilateral migration flows, I use import exposure in both the origin and the destination CZ as explanatory variables to estimate the effects on net and gross migration flows.

2.4.1 Estimates on Population and Out-Migration Rate

The main estimating equation for population and out-migration rate is as follows:

$$Y_{dt} = \alpha_0 + \beta_1 \Delta IME^{US-China}_{dt-1} + \Omega'_{dt-1}\beta_2 + \gamma_d + \gamma_t + \epsilon_{dt}$$

In this equation, Y_{dt} represents the outcome variable of interest. $\Delta IME^{US-China}_{dt-1}$ is the lagged import exposure in d from $t-1$ to t. Ω'_{dt-1} include a set of control variables that measure lagged average income, average housing cost and unemployment rate in the CZ of interest.[12] The time fixed effect γ_t teases out differences in outcome variables across time that are attributed to macroeconomic shocks common to all CZs. I include the CZ fixed effect γ_d so that the estimates for import exposure capture time-series variations within each CZ. Since the extent of import exposure is not correlated with the existing population, the cross-sectional variation is not of primary interest in this estimation. Therefore, β_1 measures the marginal effects of import exposure on CZs over time. Regressions are weighted by the CZ population share in the initial period. Robust standard errors are clustered on the CZ level to control for unobservables that are potentially serially correlated.

Estimates using population as the dependent variable are reported in Table 2.1. I compare OLS estimates in the first three columns with IV estimates in the next three columns. The IV estimate for import exposure per worker in column 4 is more than double the OLS estimate in column 1, pointing to endogeneity that leads to understatement of the true effects by OLS. The IV estimate in column 4 suggests that a one standard deviation increase in imports per worker reduces the CZ population by 1.35 percent over time, controlling

[12]Variables that measure living conditions are lagged due to potential simultaneity bias.

for other economic conditions. Since the coefficient on lagged unemployment rate is not significant, I drop this variable and present results in column 5, where estimates are almost identical to those in column 4. Therefore, housing price and average income in the previous year are positively correlated with population growth in the subsequent year. While the positive correlation between lagged income and population growth is easily rationalized by the fact that people are attracted to areas with higher incomes, it is less obvious why the correlation between lagged housing price and population growth is also positive. In column 6, where both housing price and unemployment rate are dropped from the equation, estimate for average income is the largest, implying positive correlation between average income and housing price in the previous year. This means areas with higher average incomes also have higher housing prices, and experience faster population growth. The same pattern can be observed from OLS estimates in columns 2 and 3.

Table 2.2 displays OLS and IV results using out-migration rate as the dependent variable. According to the IV estimate on import exposure per worker in column 4, a one standard deviation increase in imports per worker raises the out-migration rate by 0.28 percentage points. This estimate barely changes when dropping unemployment rate (column 5), or both the housing price and the unemployment rate (column 6) in the equation. The IV estimate in is almost double the OLS estimate shown in column 1, suggesting downward bias caused by endogeneity. Estimates on lagged average income and housing price are both positive and significant, whereas that on lagged unemployment rate is negative and insignificant. This means that higher average income and housing price in the previous year is linked to greater out-migration in the subsequent year. The positive estimate on average income implies that richer areas are more likely to send out richer migrants who are more capable of overcoming the costs of migration; the positive estimate on housing price means that more people move out of areas with faster growths in housing costs.

2.4.2 Estimates on Gross and Net Migration Flows

To gauge the effects of import exposure on bilateral migration flows, I estimate the following equation:

$$Y_{dot} = \alpha_0 + \beta_1 \Delta IME^{US-China}_{dt-1} + \beta_2 \Delta IME^{US-China}_{ot-1} + \Omega'_{dt}\beta_3 + \Omega'_{ot}\beta_4 + \gamma_{do} + \gamma_t + \epsilon_{dt}$$

The outcome variables of interest, denoted by Y_{dot}, include gross and net migration flows between d and o. To capture the effects of import exposure in both the origin and the destination, I include import exposure variables in both locations, $\Delta IME^{US-China}_{dt-1}$ and $\Delta IME^{US-China}_{ot-1}$ in the equation. Ω'_{dt} and Ω'_{ot} are sets of control variables for the origin and the destination. I exploit variations in migration flows and import exposure over time for each pair of CZ by including the CZ pair fixed effect.[13] The time fixed effect controls for migration patterns that are common to all areas in any year. Robust standard errors are clustered on the CZ pair level to account for potential serial correlation within CZ pair over time. Regressions are weighted by the sum of population shares for a CZ pair in the initial period.

Table 2.3 shows estimates using gross migration flows as the dependent variable. Greater import exposure in the origin raises gross migration flows, based on estimates across all specifications. However, estimated effects of import exposure in the destination are not all significant. Estimates on this variable are negative and significant in columns 2, 3, 5 and 6, while negative and non-significant in columns 1 and 4, where the time fixed effect and average incomes are dropped from the equation. Therefore, import exposure

[13]There is no correlation between migration flows and import exposure across different CZs, which suggests that the effects are identified within CZ over time.

in the destination does not affect migration flows as much as that in the origin, implying information asymmetry faced by migrants who are more familiar with their current residences than future ones. Consistent with previous gravity estimates, gross migration flows are positively associated with average income in the destination, and populations in both the origin and the destination, while negatively related to average income in the origin. Additionally, households move from areas with more expensive housing and higher unemployment rate to others that are the opposite.

Table 2.4 displays results for net migration flows, measured by differences in log flows between location pairs. According to estimates in column 1, a one standard deviation increase in import exposure per worker in the origin raises net outflows in the subsequent period by 0.12 log points, while the same increase in import exposure per worker in the destination lowers net inflows by 0.08 log points. The relatively smaller effect of import exposure in the destination suggested by estimates on net flows is in concert with those on gross flows. Estimates on income and population controls send a more mixed message, possibly due to the inclusion of the CZ pair fixed effect. On the other hand, the correlation between net migration and housing price or unemployment rate is consistent with gravity estimates.

2.5 Robustness

2.5.1 Using Alternative Measures of Migration Flow, Population and Income

The IRS migration data reports both the number of returns as well as the number of exemptions for migrants and county population. Returns correspond to the number of households, while exemptions approximate the number of individuals in a local area. Throughout the analyses in this paper, I use the number of returns for household count, as it is the primary unit of interest in terms of migration. To ensure results hold for the number of individuals, I replicate gravity estimations and two-stage least squares estimations in this section, using the number of individuals for both migration flows and population as dependent variables.

Additionally, the IRS county migration and income data also reports households' and migrants' income by category. I calculate the total and average wage and salary income which are used as explanatory variables. Results using alternative migration, population and income measures are shown in Tables 2.5–2.9. Estimates from the baseline gravity estimation in Table 2.5 are almost identical to the origin ones using household as the unit of analyses and gross adjusted income as income measure. The coefficients on wage and salary income are slightly bigger than those on gross adjusted income in the origin, and the reverse is true for coefficients on different income measures in the destination. Distance elasticities differ a little when using individuals as opposed to households as the dependent variable, but remain the same across alternative income measures when using the same dependent variable.

In Table 2.6, I present estimates on gross flows using alternative income and population measures for the augmented gravity equation. There are very small differences between estimates and qualitative results remain the same. The average income and wage in the origin are not correlated with gross flows, except in the last column where the origin average wage is regressed on individual flows. However, the destination income and wage exhibit strong correlation with either household or individual flows. The stark difference between coefficients on origin and destination income/wage confirms previous conclusion that the *pull* effect is strong than the *push* effect for average income.

Table 2.7 reports augmented gravity estimates on net migration flows. In addition to comparing estimates using alternative measures of income and population, I compare results using both level difference and log difference in bilateral flows. Coefficient estimates on average income/wage in the origin are negative and significant when the dependent variables are log differences, but the results are mixed using level differences. In contrast, estimates on average income/wage in the destination are positive and significant, except when net flow measured by the level difference of individuals is regressed on destination average wage, which is still positive. Distance is not correlated with net flows across all specifications, which is in concert with previous results.

Table 2.8 shows OLS and IV estimates on population, measured by the number of individuals. The preferred specification in column 3 suggests that a one standard deviation increase in imports per worker lowers the CZ population by 0.86 percent, which is smaller in magnitude than the corresponding estimate in Table 10. The IV estimates are also greater in magnitude than the OLS estimates. Results for average income and housing price are qualitatively the same as those in Table 10.

Finally, I replicate IV estimations for gross and net migration flows, using alternative measures for population and income. Results are shown in Table 2.9. The coefficient on origin import exposure is positive and significant, while that on destination import exposure is positive and insignificant. These are qualitatively the same as estimates in Table 2.4. The effect of import exposure in the origin is slightly stronger when using individual migration flows than using household flows. Results for net flows are reported in the next four columns. Despite qualitative results also being the same as in Table 2.5, there are small differences between estimates in Table 2.9 and those in Table 2.9 on the same variable. For example, import exposure in the origin has a weaker effect on net out-migration in terms of individuals versus households, but the reverse is true for import exposure in the destination.

Overall, estimates are robust to alternative measures of population and income despite small differences in magnitude.

2.5.2 Using Alternative Measures of Import Exposure

I report IV estimates using alternative measures of import exposure in Table 2.10. The first two columns include results for log population and out-migration rate on the CZ level. The next four columns include estimates on gross and net flows on the CZ pair level. There are eight different measures of import exposure to address concerns that the instrument may not capture exogenous variations in trade costs and Chinese productivities, etc. To identify the causal effect of import exposure on outcome variables, the instrument should not capture demand shocks that are correlated across developed markets. To account for potentially correlated housing demand that raise imports of related products, I construct the import exposure variable IMP1, without imports of steel, glass, cement and furniture. To ameliorate concerns on rising demand for electronic products due to technological improvements in recent decades, I drop imports of electronic equipment in IMP2. In IMP3, I exclude imports of apparel and textile as the imports of these products surged over the sample period and may have solely driven the variation in imports across developed countries.

According to Table 2.10, estimates on population using IMP1–IMP3 are negative and significant, while estimates on the out-migration rate become insignificant for IMP2 and IMP3. Results for gross and net migrant flows are qualitatively the same as benchmark estimates. It is worth noting that the magnitude of these estimates decline when some industries are excluded from the import exposure variable. The drop in magnitude is biggest for IMP3, which does not include textile and apparel imports.

In addition to dropping industries that may jeopardize the validity of the instrument, I use alternative measures of imports in variables IMP4–IMP8 to see if results hold. IMP4 uses net imports and IMP5 uses imports net of imported intermediates. IMP6 replaces imports from China with imports from low-wage countries. IMP7 uses residuals from trade gravity estimations that measure China's comparative advantage relative to the US instead of imports.[14] IMP8 substitutes import of other developed countries from China with Iranian imports from China, to address concerns on competition between US and Chinese exports in foreign markets that affect US industries.

Similar to previous comparisons, estimates on population using IMP5–IMP8 are qualitatively the same as the benchmark. Evidence on the effect of out-migration rate is mixed. However, results for gross and net migration flows are robust to alternative measures. To sum up, there are no significant differences between estimates using different measures of import exposure, except for out-migration rate.

2.6 Conclusion

This paper examines the effects of trade liberalization on population and migration flows. I develop a spatial equilibrium model that features elastic labor supply in areas that are differentiated by industry compositions. I employ two-stage least squares strategy to identify the causal effects of rising import exposure on migration, exploiting exogenous variations in lowering trade costs and rising Chinese productivities.

I find that areas subject to greater import competition experience slower population growths and more out-migration. In terms of the effects on migration flows between pairs of locations, rising import competition in the

[14]The residuals are from gravity equations where the dependent variable is relative Chinese exports versus US exports to a particular market, and the independent variables include importing country fixed effect and industry fixed effect, which capture country and industry specific characteristics of imports by a third country. Therefore, residuals capture the relative comparative advantage of China versus the US.

origin causes more out-migration, while that in the destination results in less in-migration. The effect is stronger in the origin and in the destination, suggesting possible information asymmetry faced by potential migrants who are more familiar with their current residences than future ones.

Results on the mobility response suggest labor mobility attenuates negative labor demand shocks in local labor markets. People who move because of worsening labor market conditions may find better employment opportunities in other areas. However, increased supply of workers from areas that are directly affected may further suppress wage and employment in other areas through migration. Therefore, labor mobility dilutes the effects of import competition that are initially concentrated in directly affected areas. Future work on the indirect effects may facilitate a more accurate assessment on the aggregate impacts of rising import competition from China and other low-wage countries.

2.7 Tables

TABLE 2.1: OLS and IV estimates of Import Competition on Population

	OLS			IV		
	1	2	3	4	5	6
Dependent Variable: Log Population						
Lagged IMP	-0.0317***	-0.0330***	-0.0217***	-0.0674***	-0.0684***	-0.0404***
	(0.00467)	(0.00454)	(0.00449)	(0.00903)	(0.00866)	(0.00849)
Lagged Avg Income	0.234***	0.245***	0.279***	0.217***	0.221***	0.270***
	(0.0215)	(0.0216)	(0.0237)	(0.0228)	(0.0233)	(0.0242)
Lagged HPI	0.161***	0.159***		0.157***	0.156***	
	(0.0130)	(0.0129)		(0.0129)	(0.0129)	
Lagged UR	-0.222***			-0.0883		
	(0.0555)			(0.0635)		
First Stage:						
Lagged IV				0.636***	0.645***	0.645***
				(0.0893)	(0.0876)	(0.0873)
F statistic				914.38	993.05	988.01
CZ FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
N	8,793	8,793	8,793	8,779	8,779	8,779

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Regressions are weighted by CZ's population share in the initial period. Robust standard errors, reported in parentheses, are clustered on the CZ level to account for potentially serially correlated errors. IMP and ΔIME (in the empirical strategy section) are used interchangeably in this paper. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics.

TABLE 2.2: OLS and IV estimates of Import Competition on Out-migration Rate

	OLS			IV		
	1	2	3	4	5	6
Dependent Variable: Out-migration Rate						
Lagged IMP	0.00850*	0.00848*	0.00839*	0.0141*	0.0139*	0.0139*
	(0.00492)	(0.00479)	(0.00477)	(0.00801)	(0.00783)	(0.00780)
Lagged Avg Income	0.0144***	0.0145***	0.0230***	0.0177***	0.0185***	0.0272*
	(0.00393)	(0.00440)	(0.00413)	(0.00534)	(0.00614)	(0.00597)
Lagged HPI	0.00824***	0.00825***		0.00830***	0.00844***	
	(0.00129)	(0.00128)		(0.00144)	(0.00143)	
Lagged UR	-0.00232			-0.0210		
	(0.0274)			(0.0338)		
First Stage:						
Lagged IV				0.636***	0.645***	0.645***
				(0.0893)	(0.0876)	(0.0873)
F statistic				914.38	993.05	988.01
CZ FE	Y	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
N	8,793	8,793	8,793	8,779	8,779	8,779

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Regressions are weighted by CZ's population share in the initial period. Robust standard errors, reported in parentheses, are clustered on the CZ level to account for potentially serially correlated errors. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics.

TABLE 2.3: IV estimates of Import Competition on Gross Migrant Flows

	1	2	3	4	5	6
Dependent Variable: Gross Migrant Flow						
Lagged Origin IMP	0.115***	0.0789***	0.0809***	0.124***	0.134***	0.126***
	(0.0130)	(0.0117)	(0.0128)	(0.0130)	(0.0134)	(0.0127)
Lagged Dest IMP	-0.00846	-0.0481***	-0.0565***	-0.0167	-0.0344***	-0.0733***
	(0.0113)	(0.0107)	(0.0113)	(0.0114)	(0.0112)	(0.0111)
Log Origin Population	0.270***	0.247***		0.243***	0.290***	
	(0.00923)	(0.00915)		(0.00913)	(0.00898)	
Log Dest Population	0.368***	0.351***		0.400***	0.327***	
	(0.00868)	(0.00858)		(0.00869)	(0.00864)	
Log Origin Avg Income	-0.616***	-0.670***	-0.536***		-0.296***	
	(0.0170)	(0.0162)	(0.0169)		(0.0148)	
Log Dest Avg Income	0.606***	0.562***	0.730***		0.418***	
	(0.0169)	(0.0158)	(0.0168)		(0.0148)	
Log Origin HPI	0.376***	0.350***	0.394***	0.280***		
	(0.00719)	(0.00680)	(0.00731)	(0.00632)		
Log Dest HPI	-0.283***	-0.312***	-0.262***	-0.189***		
	(0.00699)	(0.00653)	(0.00712)	(0.00612)		
Origin UR	2.529***	2.112***	2.536***	2.824***		
	(0.0659)	(0.0627)	(0.0660)	(0.0649)		
Dest UR	-2.839***	-3.282***	-2.599***	-3.126***		
	(0.0710)	(0.0678)	(0.0703)	(0.0698)		
Time FE	Y	N	Y	Y	Y	Y
CZ Pair FE	Y	Y	Y	Y	Y	Y
N	316,242	316,242	316,242	316,242	322,360	322,360

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Regressions are weighted by the sum of CZ pair populations in the initial period. Robust standard errors, reported in parentheses, are clustered on the CZ pair level to account for potentially serially correlated errors within a pair of CZs over time. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics.

TABLE 2.4: IV estimates of Import Competition on Net Migrant Flows

	1	2	3	4	5	6
Dependent Variable: Net Migrant Flow						
Lagged Origin IMP	0.574***	0.529***	0.618***	0.575***	0.563***	0.616***
	(0.0291)	(0.0244)	(0.0289)	(0.0287)	(0.0285)	(0.0282)
Lagged Dest IMP	-0.418***	-0.468***	-0.374***	-0.418***	-0.398***	-0.378***
	(0.0313)	(0.0263)	(0.0310)	(0.0309)	(0.0300)	(0.0296)
Log Origin Population	-0.380***	-0.368***		-0.361***	-0.437***	
	(0.0225)	(0.0225)		(0.0215)	(0.0223)	
Log Dest Population	-0.302***	-0.290***		-0.305***	-0.398***	
	(0.0224)	(0.0225)		(0.0217)	(0.0222)	
Log Origin Avg Income	0.120***	0.201***	-0.241***		-0.190***	
	(0.0375)	(0.0309)	(0.0336)		(0.0274)	
Log Dest Avg Income	-0.130***	-0.0475	-0.00320		0.133***	
	(0.0363)	(0.0291)	(0.0334)		(0.0259)	
Log Origin HPI	0.517***	0.522***	0.491***	0.520***		
	(0.0270)	(0.0248)	(0.0269)	(0.0241)		
Log Dest HPI	-0.531***	-0.523***	-0.542***	-0.541***		
	(0.0272)	(0.0248)	(0.0272)	(0.0241)		
Origin UR	5.136***	5.202***	4.614***	4.775***		
	(0.224)	(0.205)	(0.221)	(0.210)		
Dest UR	-3.787***	-3.726***	-3.506***	-3.396***		
	(0.224)	(0.204)	(0.220)	(0.210)		
Time FE	Y	N	Y	Y	Y	Y
CZ Pair FE	Y	Y	Y	Y	Y	Y
N	188,638	188,638	188,638	188,638	192,581	192,581

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Regressions are weighted by the sum of CZ pair populations in the initial period. Robust standard errors, reported in parentheses, are clustered on the CZ pair level to account for potentially serially correlated errors within a pair of CZs over time. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics.

TABLE 2.5: Gravity estimates using alternative population and income measures

Dependent Variable	HHD	IND	HHD	IND
Log Origin Total Income	0.151***	0.150***		
	(0.0177)	(0.0185)		
Log Dest Total Income	0.344***	0.358***		
	(0.0179)	(0.0185)		
Log Distance	-1.355***	-1.342***	-1.355***	-1.342***
	(0.00306)	(0.00318)	(0.00306)	(0.00317)
Log Origin Wage and Salary			0.171***	0.171***
			(0.0181)	(0.0189)
Log Dest Wage and Salary			0.341***	0.356***
			(0.0183)	(0.0190)
N	321,773	321,773	321,773	321,773
Adjusted R sq.	0.600	0.571	0.600	0.571

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. HHD and IND refer to migration flow measured by the number of households and individuals respectively. Wage and Salary income are reported as a subcategory of household income in IRS migration data. Robust standard errors are reported in parentheses. Origin, destination, region pair and time fixed effects are included in estimations across all columns.

TABLE 2.6: Augmented gravity estimates using alternative population and income measures – Gross Flows

Dependent Variable	HHD	IND	HHD	IND
Log Origin HHD	0.168***	0.159***		
	(0.0212)	(0.0221)		
Log Origin Avg Income	-0.0614	-0.0689		
	(0.0512)	(0.0537)		
Log Dest HHD	0.318***	0.351***		
	(0.0205)	(0.0212)		
Log Dest Avg Income	0.548***	0.597***		
	(0.0510)	(0.0532)		
Log Distance	-1.362***	-1.349***	-1.361***	-1.348***
	(0.00307)	(0.00318)	(0.00307)	(0.00318)
Log Origin IND			0.153***	0.161***
			(0.0202)	(0.0211)
Log Origin Avg Wage			0.00950	-0.100*
			(0.0540)	(0.0566)
Log Dest IND			0.347***	0.381***
			(0.0204)	(0.0211)
Log Dest Avg Wage			0.502***	0.482***
			(0.0538)	(0.0562)
N	315,655	315,655	315,655	315,655
Adjusted R sq.	0.601	0.573	0.601	0.573

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. HHD and IND refer to migration flow measured by the number of households and individuals respectively. Wage and Salary income are reported as a subcategory of household income in IRS migration data. Robust standard errors are reported in parentheses. Origin, destination, region pair and time fixed effects are included in estimations across all columns.

TABLE 2.7: Augmented gravity estimates using alternative population and income measures – Net Flows

	Level Diff		Log Diff		Level Diff		Log Diff	
Dependent Variable	HHD	IND	HHD	IND	HHD	IND	HHD	IND
Log Origin HHD	-37.16***	-115.2**	-0.269***	-0.0383				
	(10.27)	(50.91)	(0.0384)	(0.0358)				
Log Origin Avg Income	-36.62	2.023	-2.274***	-1.370***				
	(23.84)	(137.9)	(0.0958)	(0.0810)				
Log Dest HHD	53.70***	211.7**	-0.0467	0.122***				
	(15.12)	(88.66)	(0.0374)	(0.0345)				
Log Dest Avg Income	91.89***	260.6*	2.054***	1.577***				
	(25.11)	(157.1)	(0.0907)	(0.0810)				
Log Distance	-0.991	9.867	-0.991	-0.00337	-0.932	9.882	0.710	-0.00319
	(2.386)	(10.82)	(2.386)	(0.00373)	(2.386)	(10.83)	(1.476)	(0.00373)
Log Origin IND					-39.11***	-121.5**	-0.388***	-0.117***
					(9.897)	(49.07)	(0.0371)	(0.0347)
Log Origin Avg Wage					-14.50	53.80	-1.878***	-1.275***
					(22.12)	(125.9)	(0.0990)	(0.0843)
Log Dest IND					58.58***	225.1**	0.0624*	0.206***
					(15.17)	(89.10)	(0.0373)	(0.0350)
Log Dest Avg Wage					58.02**	176.5	1.771***	1.443***
					(23.39)	(143.9)	(0.0955)	(0.0841)
N	188,268	188,268	188,268	188,268	188,268	188,268	188,268	188,268
Adjusted R sq.	0.069	0.077	0.378	0.371	0.069	0.077	0.377	0.369

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. HHD and IND refer to migration flow measured by the number of households and individuals respectively. Wage and Salary income are reported as a subcategory of household income in IRS migration data. Robust standard errors are reported in parentheses. Origin, destination, region pair and time fixed effects are included in estimations across all columns.

TABLE 2.8: The effects of import exposure on population, measure by the number of individuals

	OLS		IV	
	1	2	3	4
Dependent Variable: Log Population				
Lagged IMP	-0.0176***	-0.0177***	-0.0436***	-0.0426***
	(0.00529)	(0.00519)	(0.00994)	(0.00959)
Lagged Avg Income	0.259***	0.260***	0.247***	0.243***
	(0.0230)	(0.0229)	(0.0241)	(0.0246)
Lagged HPI	0.160***	0.160***	0.158***	0.158***
	(0.0133)	(0.0132)	(0.0133)	(0.0132)
Lagged UR	-0.0215		0.0787	
	(0.0588)		(0.0679)	
CZ FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
N	8,793	8,793	8,779	8,779

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Regressions are weighted by CZ's population share in the initial period. Robust standard errors, reported in parentheses, are clustered on the CZ level to account for potentially serially correlated errors. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics.

TABLE 2.9: IV estimates on gross and net individual flows

Dependent Variable:	Gross Individual Flows				Net Individual Flows			
Lagged Origin IMP	0.131***	0.0928***	0.156***	0.149***	0.388***	0.417***	0.385***	0.377***
	(0.0156)	(0.0146)	(0.0156)	(0.0143)	(0.0409)	(0.0335)	(0.0410)	(0.0403)
Lagged Dest IMP	0.0134	-0.0295**	-0.0135	-0.0629***	-0.488***	-0.457***	-0.497***	-0.487***
	(0.0125)	(0.0126)	(0.0121)	(0.0118)	(0.0412)	(0.0328)	(0.0411)	(0.0406)
Log Origin IND	0.231***	0.219***	0.285***		0.0423*	0.0426*		
	(0.0150)	(0.0150)	(0.0135)		(0.0241)	(0.0241)		
Log Dest IND	0.434***	0.423***	0.380***		0.0275	0.0278		
	(0.0131)	(0.0130)	(0.0133)		(0.0242)	(0.0242)		
Log Origin Avg Wage	-0.620***	-0.728***	-0.295***		0.240***	0.155***		
	(0.0276)	(0.0276)	(0.0241)		(0.0591)	(0.0425)		
Log Dest Avg Wage	0.504***	0.398***	0.312***		0.00593	-0.0794*		
	(0.0273)	(0.0267)	(0.0239)		(0.0594)	(0.0423)		
Log Origin HPI	0.406***	0.408***			0.414***	0.444***	0.459***	
	(0.0116)	(0.0111)			(0.0355)	(0.0330)	(0.0342)	
Log Dest HPI	-0.306***	-0.304***			-0.520***	-0.490***	-0.514***	
	(0.0112)	(0.0102)			(0.0352)	(0.0328)	(0.0337)	
Origin UR	2.670***	2.177***			4.041***	4.244***	3.471***	
	(0.110)	(0.101)			(0.299)	(0.286)	(0.276)	
Dest UR	-2.824***	-3.342***			-3.932***	-3.724***	-3.463***	
	(0.134)	(0.120)			(0.296)	(0.282)	(0.271)	
Log Dist	-0.000360***	-0.000405***			0.00118***	0.00117***	0.00100**	
	(0.0000849)	(0.0000790)			(0.000409)	(0.000408)	(0.000394)	
Time FE	Y	N	Y	Y	Y	N	Y	Y
N	316,242	316,242	322,360	322,360	188,268	188,268	188,268	188,268

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Regressions are weighted by the sum of CZ pair's population share in the initial period. The CZ pair fixed effect is included in all columns. Robust standard errors, reported in parentheses, are clustered on the CZ pair level to account for potentially serially correlated errors. HPI stands for House Price Index, which comes from the Federal Housing and Finance Agency. UR refers to Unemployment rate, which is calculated from Local Area Unemployment Statistics provided by the Bureau of Labor Statistics.

TABLE 2.10: IV estimates on population, out-migration rate and migration flows using alternative import measures

			Gross Flow		Net Flow	
Dependent Variable	POP	OMR	Origin	Destination	Origin	Destination
Lagged IMP1	-0.0674***	0.0141*	0.106***	-0.00923	0.531***	-0.421***
	(0.00903)	(0.00801)	(0.0125)	(0.0110)	(0.0226)	(0.0335)
Lagged IMP2	-0.0691***	0.0139*	0.112***	-0.00811	0.524***	-0.408***
	(0.00919)	(0.00787)	(0.0132)	(0.0093)	(0.0194)	(0.0286)
Lagged IMP3	-0.0400***	0.0137*	0.0854***	-0.00797	0.417***	-0.385***
	(0.00937)	(0.00808)	(0.0120)	(0.0118)	(0.0230)	(0.0292)
Lagged IMP4	-0.0652***	0.0142*	0.102***	-0.0162	0.586***	-0.457***
	(0.00913)	(0.00806)	(0.0152)	(0.0132)	(0.0286)	(0.0329)
Lagged IMP5	-0.0669***	0.0140*	0.0921***	-0.00941	0.566***	-0.433***
	(0.00905)	(0.00796)	(0.0183)	(0.0111)	(0.0249)	(0.0315)
Lagged IMP6	-0.0560***	0.0121	0.0867***	-0.00698	0.342***	-0.316***
	(0.00954)	(0.00754)	(0.0291)	(0.0175)	(0.0297)	(0.0405)
Lagged IMP7	-0.0581***	0.0132	0.0518*	-0.00726	0.454***	-0.387***
	(0.0117)	(0.00921)	(0.0302)	(0.0291)	(0.0491)	(0.0308)
Lagged IMP8	-0.0711***	0.0149*	0.132***	-0.0157*	0.638***	-0.532***
	(0.00962)	(0.00863)	(0.0215)	(0.00923)	(0.0220)	(0.0288)
CZ FE	Y	Y	N	N	N	N
CZ Pair FE	N	N	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y	Y
N	8,779	8,779	316,242	316,242	188,638	188,638

Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. POP refers population, and OMR refers to Out-Migration Rate. IMP1–IMP8 are alternative measures for import exposure. Detailed discussions of these variables are in Section 7. Regressions are weighted by the CZ population for POP and OMR, and sum of CZ pair populations for gross and net flows in the initial period. Robust standard errors, reported in parentheses, are clustered on the CZ level for POP and OMR, and CZ pair level for gross and net flows.

Chapter 3

The Labor Market Effects of Offshoring and Immigration in US Service Sectors

3.1 Introduction

The opposing effects of globalization on the US labor market have stirred heated debates on trade and immigration policies. Advocates for more restrictive policies argue that native workers are displaced by immigrants who possess similar skills, and that the jobs performed by native workers are offshored to countries with lower labor costs. On the other hand, an increase in the productivity of domestic firms, generated by easier immigration and offshoring, is associated with higher levels of native employment. For example, high-skilled immigrants contribute to innovations which raise firm's productivity and competitiveness, resulting in the expansion of domestic employment. At the same time US firms, especially multinationals, can provide more job opportunities or higher pay for native workers by cost savings through offshoring to low-cost countries. Also, such productivity effect and displacement effect are likely to differ by industries, since the relative skills of native, immigrant or offshore workers, which vary across different industries, affect the level of substitutability between them. This paper explores the skill distribution of these groups of workers, and examines the productivity and substitution effect of immigration and offshoring in service sectors.

This paper is closely related to Ottaviano, Peri, and Wright (2013) (henceforth OPW), who study the effects of immigration and offshoring on US manufacturing industries. Building on Grossman and Rossi-Hansberg (2008), OPW develop a partial equilibrium model of task assignment among heterogeneous native, immigrant and offshore workers. They find that a reduction in offshoring costs decreases the employment shares of both native and immigrant workers. However, only the employment share of offshore workers is affected with a reduction in immigration costs. On the other hand, offshoring does not have any significant productivity effect on natives while immigration has a positive and mildly significant effect in terms of employment levels. In addition, they find that increased offshoring raises the average task complexity of native tasks, widening the gap between native and immigrant task complexity, whereas increased immigration has no effect on the average complexity of native tasks. Based on these observations, they conclude that native and immigrant workers are concentrated at the opposite ends of the task complexity spectrum, and offshore workers specialize in tasks of intermediate complexity.

While the changing dynamics of US manufacturing due to immigration and offshoring have received much attention (see OPW (2013) and Peri and Sparber (2009) for example), relatively little work has looked at how service sector employment may be affected by trade liberalization or immigration policies, even though most US workers are employed in this sector. I study the effects of immigration and offshoring for US workers using data in the service sectors. These effects are likely to be different in services as compared to manufacturing, since the pattern of immigration and offshoring may vary both within service sectors and between service and manufacturing. For example, there is likely to be greater demand for high-skilled immigrants in computer system design and related services than in restaurants and food services or any industry within manufacturing. Such differences are potentially important in determining the productivity and substitution effects of offshoring and immigration on native workers.

In order to measure the productivity and substitution effect of immigration and offshoring in service sectors, I use employment and education data on immigrants and natives from the American Community Survey (ACS) and employment data on offshore workers from the Bureau of Economic analysis' multinationals dataset. I match the occupation data from the ACS with the skill indices from the US Department of Labor's O*NET abilities survey, so that I can observe the responses of native and immigrant task complexity to changes in the costs of immigration and offshoring. By comparing the education attainment and average skills of the tasks (or "occupations") for native and immigrant workers, I find that immigrants perform tasks that are more intensive in manual skills than native workers, despite the differentiated distributions of education levels between these two groups in each sector.

Next, I gauge evidence on the skill complexity of offshore workers based on the substitution patterns between natives, immigrants and offshore workers, since the skills of tasks performed by offshore workers are unobserved. Based on results from 2SLS regressions employing cost variables of immigration and offshoring as instruments, I find no substitution between native and immigrant workers, and modest substitution between native and offshore workers. While these observations may suggest that the skills of natives and offshore workers are more similar than those of natives and immigrants, further evidence on task upgrading implies this is not necessarily true, since offshore workers may only compete with low-skilled native workers. Meanwhile, I find positive productivity effect of immigration and negative productivity effect of offshoring on native employment levels. A reduction in the cost of offshoring is associated with a decrease in the share of native workers, while the native share increases with easier immigration.

Additionally, both native and immigrant workers upgrade their skill complexity when faced with increased offshoring, and the gap between native and immigrant complexity narrows. These results are consistent with a skill distribution in which offshore workers are at the lower end of the task complexity spectrum, whereas native workers are at the higher end and immigrants specialize in tasks of intermediate complexity. The substitution patterns suggest that the influx of immigrants during the sample period were largely high-skilled ones who complement the skills of native workers and therefore generates a positive productivity effect. At the same time, the offshored tasks were more intensive in manual skills, which reduced the employment of native workers, especially those performing low-skilled tasks.

The rest of this paper is organized in four sections. The next section reviews relevant literature on the labor market effects of immigration and offshoring. Section 3 presents data descriptions of native and immigrants' education, employment share, etc. Section 4 describes the empirical strategy and econometric results. Section 5 concludes.

3.2 Literature Review

This paper is related to several strands of literature that study the impact of immigration and/or offshoring on US labor markets. The first strand focuses on the effect of offshoring and how the skill distribution of native workers may be affected by the types of tasks that are offshored. Grossman and Rossi-Hansberg (2008) develop a simple theory of offshoring which identifies a productivity effect that is brought by reductions in the cost of offshoring. In their story, offshoring displaces workers but increases domestic wages because of cost savings in production. However, the wage effects of offshoring depend on the crucial assumption that domestic employment doesn't change when tasks are relocated abroad.

Empirically, Crino (2010) study the impact of service offshoring on white-collar employment, and find positive employment effect of offshoring on high-skilled occupations, while it lowers employment for medium or low-skill ones. Hummels et al. (2014) also document positive employment and wage effects of offshoring on skilled workers, using firm-level data for Danish workers. Similar to the findings in Crino (2010), low-skilled workers suffer from employment and wage losses due to offshoring. Wright (2014) uses data on US manufacturing sector and concludes that offshoring affects employment of production workers negatively, while expends non-production employment slightly. Harrison and McMillan (2011) find that the substitution between native and offshore employment in manufacturing is higher for multinationals that perform more similar tasks home and abroad, despite the imprecise measurement of offshore employment which is discussed in more detail below.

Another strand of literature analyzes the effect of immigration on US labor markets. More specifically, high-skilled immigration has positive employment effect on skilled workers in the US, as is documented in Kerr and Lincoln (2010). Similar evidence has been found by Peri, Shih, and Sparber (2015), who look at the impact of foreign STEM (Science, Technology, Engineering and Mathematics) workers on native wage and employment. Although high-skilled immigration tends to generate productivity effect on domestic labor market for skilled workers, earlier research have shown differential impact on low-skilled workers. For example, Borjas (2003) and Borjas and Katz (2007) find negative employment and wage effects of immigration on low-skilled native workers.

The third and more sparse strand of literature integrates the analyses of immigration and offshoring on domestic labor market outcomes. Olney (2012) presents a simple mode that examines the impact of offshoring and immigration on wages and tests these predictions using state-industry-year panel data. According to his empirical analysis, the productivity effect causes offshoring to have a larger positive impact on low-skilled wages than immigration, but this gap decreases with the workers' skill level. Ottaviano, Peri, and Wright (2013) analyze the impact of immigration and offshoring on US labor market in a joint framework. They develop a theoretical model adapted from Grossman and Rossi-Hansberg (2008), and add immigration on top of offshoring to the original model. By using US employment data from 58 manufacturing industries over the years 2000-2007, they test the predictions from their extended model and conclude that offshoring does not harm native employment since it generates productivity effect that offsets the substitution effect. Furthermore, immigration benefits low-skilled native workers because the positive productivity effect dominates the substitution effect that is mitigated by "task upgrading" of native workers.

Since direct measures of "offshore employment" are unobserved in the data, OPW use foreign affiliate employment of US Multinationals as a proxy for the offshore employment variable. The use of this proxy, however, relies on two key assumptions that are questionable. First, all tasks performed in foreign affiliates are to be carried out in the US, if not offshored. In other words, all workers at foreign affiliates serve the purpose of supplying tasks to US parent companies. This assumption is at odds with the reality that US Multinationals set up foreign affiliates, in part, to serve foreign markets. This means some workers at foreign affiliates serve the foreign market instead of supply tasks to US headquarters. Therefore, the authors overestimate "offshore employment" at foreign affiliates, by using foreign affiliate employment directly.

The second assumption is that the labor content per unit of subcontracted intermediate inputs (produced by unaffiliated foreign firms) is the same as for production in US affiliates in the same industry. In the published version of the paper, the authors use total imports (including those from foreign affiliates and unaffiliated foreign subcontractors) and foreign affiliate employment for their estimation. Under this second assumption, this strategy captures the variations in imports and "offshore employment' associated with these imports. However, the labor content per unit of production by unaffiliated foreign subcontractors may differ substantially from that by foreign affiliates in some industries. Thus the measurement of "offshore employment" associated with imports is potentially flawed. Although strong assumptions are imposed in measuring offshore employment, there is hardly a better alternative method, given the nature of this variable and the availability of the data. Therefore, I use a similar method as in OPW to measure offshore employment.

Besides the limitations in measurement, OPW's empirical specification for estimating the substitution between native, immigrant and offshore workers is problematic. Although they regress native employment share on immigrant and offshore employment shares separately, the regression equation includes both shares on the right hand side. Under this specification, if they obtain insignificant estimate for the immigrant share, then the estimate for the offshore employment share has to be negative and significant. This is because all three shares sum up to 1. I correct this specification in my paper.

3.3 Data and Descriptive Evidence

3.3.1 Data on natives and immigrants from the Census-American Community Survey and the Bureau of Economic Analysis' multinationals database

The main datasets I use to create the employment variables for native, immigrant and offshore workers are the IPUMS samples of the Census-American Community Survey and the Direct Investment and Multinational Companies dataset from the Bureau of Economic Analysis. The IPUMS sample of the Census-ACS contains information on the birthplace, educational level, employment status, occupation, industry, etc of individual workers in the US. The Direct Investment and Multinationals Companies dataset includes restricted 2-digit and 4-digit NAICS industry level employment and wage information of nonbank foreign affiliates of US parent companies for the period 1999-2008. I extract data on native and immigrant workers from the IPUMS sample and on offshore workers from the BEA's multinationals dataset. In addition to these two datasets, I exploit information on the skill levels of native and immigrant workers by matching the occupation of each individual in the IPUMS sample with that in the O*NET abilities survey from the US Department of Labor described in the next subsection. Matching the two datasets provides information on the relative skill levels of natives as compared to immigrants, which have implications on the substitutability between these two groups of workers.

To obtain the industry-level employment and average wage of native and immigrant workers, I aggregate individual-level data into major (2-digit NAICS) service sectors from the IPUMS sample of the Census-ACS for the period 2000-2009. When aggregating individuals, I only include workers not living in group quarters and who work at least one week during the year. I also weight each worker by the sample weights assigned by the ACS to make the sample representative on the national level. I define "immigrants" as workers who were born in any non-US territory, and "native" workers as rest of the workers in the sample. Unlike the IPUMS samples with few undisclosed information, the BEA's multinationals dataset has many blanked-out columns on the 4-digit NAICS industry level for confidentiality reasons, therefore I only include offshore workers' employment data on major (2-digit NAICS) service sectors which are mostly available. And since the Census-ACS data follows the IND classification of industry, which differs slightly from the NAICS industry classification in the BEA's multinationals dataset, I matched the two datasets using the concordance table A1 in the appendix.

3.3.2 Data on the skills of occupations from the US Department of Labor's O*NET abilities database

The information on the skill levels of native and immigrant workers comes from the occupational characteristics data by the US Department of Labor's (2012) O*NET abilities survey. This survey is based on the Standard Occupation Classification (SOC) and includes numerical values that describe the importance of different skills required in each occupation. To generate the skill indices based on these values, I use the method by Peri and Sparber, 2009 and merge these task-specific values with individual workers in the 2000 Census, and calculate the percentile score of each skill for a given occupation (an implicit assumption here is that the 2000 Census is collectively representative of the US workforce). This measure will imply the relative importance of a given skill among US workers ranging between zero and one. For example, an occupation with a score of 0.95 for some skill indicates that only 5% of workers in the US are supplying this skill more intensively in the US. After generating each skill index, I create the "main" skill indices – Cognitive Intensity, Communication Intensity and Manual Intensity, as averages of the relevant skill variables.

Each skill index corresponds to a distinct set of relevant variables: Cognitive Intensity is the average of 10 variables such as Fluency of Ideas, Originality and Mathematical Reasoning; Communication Intensity is the average of 4 variables including Oral Comprehension, Written Comprehension, etc; Manual Intensity is the average of 19 variables such as Arm-Hand Steadiness, Manual Dexterity and Control Precision. In addition to the three skill indices, I create a Complexity index (following OPW) that describes the intensity of a task in cognitive and communication skills relative to manual skills. This Complexity index is defined by the following equation:

$$Complexity = ln(\frac{Cognitive + Communication}{Manual})$$

The range of this index is from negative infinity to positive infinity. By assigning the skill indices to each individual in the IPUMS sample according to his/her occupation, I analyze the relative skills of native and immigrant workers, and infer the skills of offshore workers using further information on employment.

3.3.3 Descriptive Statistics

Comparing the education levels of native and immigrant workers

Since education plays an important role in determining the relative skill levels of tasks performed by natives and immigrants, I first look at how these two groups of workers differ in their educational levels in each of the 9 service sectors reported in the concordance table. In figure 1 (included in the Graph section at the end), I plot the histograms of educational attainment in each service sector by native and immigrant status, over the period 2000-2009. This data sample includes 5,655,219 observations, of which 4,861,824 are native-born and 793,395 are foreign-born. In the histograms, the educational levels of interest are Grade 12 and 4 years of college, as they correspond to the typical years required for high school and college degree.

From the histograms in figure 3.1, the compositions of native and immigrant workers at various educational levels differ across service sectors. In sectors such as the Professional, scientific, and technical services, there's a higher proportion of immigrants who completed 4 years of college of more, whereas this proportion is higher for native workers in sectors like the Accommodation and food services. In addition to that, the left tail of the educational attainment distribution for immigrant workers is "fatter" in almost all service sectors except for Professional, scientific and technical services, as well as Management of companies and enterprises. This seems unsurprising because in industries related to data sciences and computer engineering, higher-level qualifications are required for immigrants (thus "high-skilled") to stay and work in the US, resulting in higher college completion rates. However, companies in accommodation and food services may be willing to hire immigrants equipped with less education (thus "low-skilled") to lower their cost of employment. These observations show that the composition of skills for each sector varies for natives and immigrants, resulting in different substitutability between them across service sectors.

Comparing the skills associated with tasks performed by native and immigrant workers

Next, I compare the skills of natives and immigrants, which are inferred by the occupations of each group, to obtain more evidence on the substitutability between them. Peri and Sparber (2009) suggested that native workers may move to tasks (or occupations) that are more intensive in communication and language skills in response to an influx of less-educated immigrants. Occupation switching by native workers, who have comparative advantage in communication-intensive tasks, reduces wage losses of native workers and also weakens the substitution effect of immigrants. Although they look at native and immigrant workers across all sectors, including manufacturing and service, service sectors employ most of the workers in the US. Therefore, we would expect similar patterns of skill composition between native and immigrant workers in service sectors only, which is verified in Figure 2.

Figure 3.2 looks at the sector share of immigrant employment versus skills associated with each particular sector. Each data point is an occupation-industry-year cell. It derives from a sample of 1,665 occupations in service sectors assigned with skill indices from the O*NET database, over the period of 2000-2009. And only occupations with over 5,000 workers are included in the sample. The four panels of this figure suggest that native workers perform tasks that are more intensive in cognitive and communication skills, whereas immigrant workers specialize in manual-intensive tasks. These findings are in line with the results from Peri and Sparber (2009) and Ottaviano, Peri, and Wright (2013), although the former used employment data for all sectors and the latter focused on manufacturing only. While figure 2 provides information on the overall share of employment given skill indices, we also look at how each skill (on average) changes over time for the two groups of workers, to see if there is any pattern of co-variation between skill adjustments by natives and immigrants.

Figure 3.3 plots the evolution of skill indices for native and immigrant workers over time. I calculate the weighted average of all four indices (including complexity) for native and immigrant workers. From these panels, the cognitive and communication skills of native workers remained constant, while the manual skill has increased slightly since 2006. For immigrant workers, their cognitive and communication skills declined until 2002, rebounded afterwards before dipping again in 2007, and then recovered slightly. Since 2007, immigrants have been performing more tasks intensive in cognitive and communication skills and less manual-intensive tasks, while native workers have been performing more manual-intensive tasks. This observation may be the result of the net out-migration of Mexican workers in the US during the Great Recession starting around this year. However, a closer look at how the Great Recession may associate with this observation requires more data and is beyond the scope of this paper. Although we can't observe the complexity of offshored tasks, implications may arise from the way offshore employment impacts the complexity of native and immigrant tasks. If offshore employment affects the complexity of tasks performed by one group of workers, the complexity of tasks by this group is thus more similar to that offshore workers than the other group. Figure 3.4 plots the change in the complexity of tasks performed by natives and immigrants against the change in the shares of offshore and immigrant employment, across services sectors over 2000-2004. Note that the sample period is shorter for analyses related to offshoring, since the data on service imports is available until 2004, and only 5 service sectors can be matched from the service import data.

The first two panels in Figure 3.4 suggest little evidence of native and immigrant task adjustments due to changes in offshore employment. This observation may result from the lack of variation in offshore employment changes since most data points are clustered around zero in the horizontal axis. However, increase in the share of immigrant employment seems to have slightly negative effects on the complexity of native tasks, as suggested by panel C. This result implies that the native workers are "downgrading" their skills in sectors with increasing share of immigrant employment. One possible explanation is that large influx of immigrants since 2000 may occur within service sectors that demand more high-skilled immigrants.

Overview of the employment information

In Figure 3.5, I plot the employment shares of native, immigrant and offshore workers in 5 service sectors over the period of 2000-2004. Overall, the immigrant share of employment increased slightly across all sectors over the sample period, among which the share increased the most (more than doubles) in management of companies and enterprises. Offshore employment share dropped by about 20% in this sector, while native share increased by almost the same percentage. Native and immigrant workers appear to have substituted for offshore workers since 2002 in management of companies and enterprises, although the shares remained steady in all other sectors.

While evidence from native complexity and immigrant employment suggest that there exists moderate substitutability between these two groups of workers, information on the employment shares seems to tell a slightly different story. Figure 3.6 explores the co-variation between employment shares of all three groups of workers in the 5 service sectors. Panel A in Figure 3.6 depicts the correlation between native and immigrant employment shares. Panel B contains the same information for native and offshore workers and panel C shows the employment shares of immigrant and offshore workers. This figure implies no substitution between native and immigrant workers, while there exist strong substitution between native and offshore workers and moderate substitution between immigrant and offshore workers. Although the direct conclusions from this figure are high substitutability between native and offshore workers, and medium substitutability between immigrant and offshore workers, such result may be inaccurate because of the lack of variation and the huge drop in offshore employment share in 2003, as is shown in the previous figure.

Despite the mixed evidence on the substitutability among the three groups of workers, we could nonetheless look for any productivity effect of immigration or offshoring, as well as the impact of those on the wage of native workers. In Figure 3.7, I plot yearly changes in native employment and wage against immigrant and offshore employment changes respectively. I use the same sample as in previous figures. Panel A suggests that changes in immigrant employment is negatively correlated with change in native wages, while panel B depicts positive correlation between the employment of immigrant and native workers. Offshore employment changes are not significantly correlated with change in native wages and employment, as is shown in Panel C and Panel D. This figure shows the productivity effect of immigration on native workers, while the increased supply of immigrants suppress the wage of native workers.

Based on the observations, I conclude that native workers differ from immigrant workers in terms of their average educational attainment. Native workers perform tasks that are more intensive in cognitive and communication skills, whereas immigrant workers perform more manual-intensive tasks. As a result, there is little substitution between natives and immigrants. Also, native workers "move to" tasks that are more intensive in manual skills when faced with increased immigration, suggesting that the new immigrants are relatively high-skilled. Relatively little evidence has been found for the effect of offshoring on native workers, which I investigate further in the next section.

3.4 Empirical Specifications and Econometric Results

Although descriptions of the data have provided some evidence on the potential substitution and productivity effect of immigration and offshoring, I examine these effects more systematically through the empirical analyses in this section. The analyses will be carried out in three parts: First, I estimate the effects of immigration and offshoring on native employment share, which sheds light on the substitutability between different groups of workers; Second, I estimate the effects of immigration and offshoring on total employment levels to gauge any productivity effect; Third, I test whether the distribution of task complexity coincides with my assumptions as well as with the substitution patterns between different groups of workers.

I employ instruments for the cost of immigration and offshoring in this part of the analysis. Since the cost of offshoring is unobserved in service sectors, I use the dollar amount of service imports as a proxy for the cost of offshoring, assuming that the variation in service offshoring is mostly driven by the variation in costs. The effects of immigration and offshoring on native and immigrant employment as well as task specialization will thus be identified using variations in immigration and offshoring cost variables within sectors over time. These cost variables are discussed in more detail in the paragraphs that follow.

3.4.1 Costs of Immigration and Offshoring

While most of the costs associated with trade in goods can be directly measured (distance, tariff for instance), cost of service trade may involve unobserved factors such as tradability. For example, services provided by waiters/waitresses in a restaurant are not tradable. However, the amount of imported service incorporates the information on various costs that are related to offshoring. Despite the potential endogeneity issues that arise from this instrument, it is the best estimate of offshoring costs available in the data. Therefore, I use imported services as an instrument for offshore employment.

The data on imported purchased services comes from a published table in Yuskavage, Strassner, and Medeiros (2009), and it was based on these authors' calculations from unpublished data in BEA's Annual Industry Accounts. In their paper about outsourcing and imported service in BEA's industry accounts, they described the methodologies used for BEA's Annual Industry Accounts and included various data on service imports. The main assumption they made in calculating service imports is that for each comparable commodity used by an industry, the portion attributable to imports was equal to the economy wide share of imports in the total supply of the commodity. Since their data only spans the period 1997-2004, and the IPUMS sample was surveyed annually since 2000, I could only use this proxy for the cost of offshoring from 2000 to 2004. Another limitation of this data is that those are available mostly on the 2-digit NAICS major industry level.

In order to proxy for the cost of immigration by industry and year, I generate an imputed immigrant share variable following the method in Ottaviano, Peri, and Wright (2013). This method was first proposed by Altonji and Card (1991) and Card (2001) to facilitate the identification of cost-driven shifts in immigration. This variable exploits differences in the presence of immigrant groups (from different countries) across industries, based on the fact that changes in the presence of foreigners are driven by the cost of migrating and domestic conditions specific to their countries of origin. To create this index, I first compute the share of immigrant workers by countries of origin in Year 2000, and then augment it by the aggregate growth rate of the specific immigrant group's population in the US relative to the total US population. The imputed share of immigrants in total employment is then the sum of immigrant shares over origin-groups within each sector.

3.4.2 Effect on Employment Shares and Employment Levels

The main estimating equations for the substitutability of native, immigrant and offshore workers using costs of immigration and offshoring as instruments are as follow:

$$NS_{it} = \beta_{OD} * OS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.1}$$

$$NS_{it} = \beta_{MD} * MS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.2}$$

$$MS_{it} = \beta_{OM} * OS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.3}$$

$$OS_{it} = \beta_{MO} * MS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.4}$$

In these equations, NS_{it}, MS_{it} and OS_{it} represent the employment shares of native, immigrant and offshore workers. Imputed immigrant share is denoted as IMS_{it} and serves as the instrument for the share of immigrant employment. Share of offshore workers is instrumented by imported purchased services SI_{it}. Industry and year fixed effects are denoted as ψ_i and ψ_t respectively, where x represents the categories O, M, D for offshore, immigrant and native. In equations (3.1) and (3.2), I estimate how changes in the employment shares of immigrant and offshore workers affect native employment share separately, using instruments that proxy costs. By controlling for sector and year, I tease out time-invariant differences between sectors and the common trend across sectors over time, so that the equations only exploit within industry variations of immigration and offshoring costs.

In equations (3.3) and (3.4), I explore the substitutability between immigrant and offshore workers, using the same instruments for the employment shares as in the first equation. The coefficients in equations (3.1) and (3.2) (second stage equation) are interpreted as the percentage change in the native share of employment associated with 1 percent change in offshore share and immigrant share respectively. Coefficients for equations (3.3) and (3.4) represent the percentage change in immigrant/offshore share in response to 1 percent change in offshore/immigrant share. If any coefficient is estimated to be negative and significant, then its corresponding employment group substitutes for the group represented by the dependent variable. However, an estimated coefficient between -1 and 0 implies that, if the independent variable increases by 1 percent, then the employment level for the same group as the dependent variable increases, despite its decreasing employment share.

In Table 3.1, I report the estimated coefficients and Heteroskedasticity-robust standard errors for the first 3 equations. Estimates from equations (3.1) and (3.2) are recorded in the first two rows of this table, and estimates from equation (3.3) and (3.4) are included in Row 5 and Row 7 respectively. As in OPW (2013), I also regress native employment share directly on imported purchased services and imputed immigrant share, and the coefficient estimates are reported in Row 3 and Row 4. Finally, Row 6 and Row 8 include estimates from regressions of immigrant share on service imports and offshore share on imputed immigrant share. According to the estimates, there is moderate substitutability between native and offshore workers, as well as between offshore and immigrant workers, although the magnitude of substitution is lower for the latter. Coefficients on immigrant share using imputed immigrant share as an instrument seems erratic, and the results seem to suffer from the lack of variation in the data as standard errors are usually large.

Next, I explore potential "productivity effect" by estimating the impact of immigration and offshoring on native employment level, as well as total employment. If there exists positive and significant productivity effect of immigration or offshoring, then I expect total employment to move in the opposite direction with immigration or offshoring costs. And if the productivity effect dominates the substitution effect, then native employment is expected to increase, even though offshore workers may substitute native ones (as shown in the Table 3.1). The empirical specifications here are similar to equations (3.1)-(3.4), with the following exceptions: 1. Employment shares are replaced by employment levels; 2. Total employment level is added as a dependent variable:

$$NL_{it} = \beta_{ON} * OL_{it} + \beta_{MN} * ML_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.5}$$

$$ML_{it} = \beta_{OM} * OL_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.6}$$

$$OL_{it} = \beta_{MO} * ML_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.7}$$

$$TL_{it} = \beta_{OT} * OL_{it} + \beta_{MT} * ML_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.8}$$

Again, industry and time fixed effects are denoted as ψ_i and ψ_t , and ϵ_{it} s are potentially serially correlated errors. Employment levels of native, immigrant and offshore workers are represented by NL_{it}, ML_{it} and OL_{it}. In equation (3.8), the dependent variable TL_{it} is the total employment of all three groups of workers. I expect the sign of β_{OT} to be positive if the overall productivity effect is positive, and the signs of β_{ON} and β_{MN} to also be positive if the productivity effect of immigration and offshoring outweighs the respective substitution effect. Table 3.2 records regression estimates for equations (3.5) through (3.7). Similar to Table 3.1, the first two columns record estimates from the 2SLS regression on native employment. The positive coefficient of the immigration employment variable suggests positive productivity effect on native employment. Although the coefficient on the offshore employment variable appears negative, the standard error is also large. Therefore, no evidence of productivity effect (on native employment) for offshoring is found. Estimates for the effect of immigration and offshoring on total employment levels (thus equation 3.8) follow similar patterns, and are reported in Table 3.3.

3.4.3 Effect on Skills

According to the analyses on employment levels and shares of the different worker groups in this section, immigration benefits native employment while offshoring effects it negatively. Findings from the regression analyses are consistent with the evidence from the data description. Next, I look at how native workers may adjust their tasks when they are faced with increased level of immigration and offshoring, through OLS and 2SLS regressions for each of their skill index. The empirical specifications are listed below:

$$NI_{yit} = \beta_{ON} * OS_{it} + \beta_{MN} * MS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.9}$$

$$MI_{yit} = \beta_{OM} * OS_{it} + \beta_{MI} * MS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.10}$$

$$GI_{it} = \beta_{OG} * OS_{it} + \beta_{MG} * MS_{it} + \psi_i + \psi_t + \epsilon_{it} \tag{3.11}$$

Equations (3.9) and (3.10) examine the impact of immigration and offshoring on the skill indices of native and immigrant workers. Since the data on skill indices for offshore workers are unavailable, I try to gauge evidence on the relative level of skills for offshore workers compared to the other two groups. The skill indices for natives are denoted as NI_{yit}, where the subscript y represents each category of the skill–cognitive, communication, manual and complexity. Similarly, immigrant's average skill is denoted as MI_{yit}. Equation (3.10) analyzes how the gap between native and immigrant skill varies by offshoring, so the dependent variable GI_{it} is the difference between native and immigrant complexity, or $(NI_{it} - MI_{it})$. I report estimates for these regressions in Table 3.4. According to estimates reported in the table, offshoring is positively correlated with native complexity. Combining this evidence with the substitution and productivity effect from previous regressions, native workers seem to "upgrade" their skill when faced with competition from offshore workers. In the direct OLS regression where the dependent variables are immigrant complexity and the difference between native and immigrant complexity, I find that offshoring pushes up immigrant complexity, and narrows the gap between native and immigrant complexity. These results imply that offshore workers may perform less "complex" tasks than natives and immigrants, since increased offshoring pushes both native and immigrants up the skill distribution. Additionally, native workers upgrade their skills more than immigrants. However, a direct consequence of this skill distribution, where immigrants are located between native and offshore workers, would be higher extent of substitution between natives and immigrants, as compared to between native and offshore workers. Such a phenomenon can be explained by two facts: 1. the influx of immigrants over the sample period was mostly higher-skilled which complement native skills; 2. the tasks offshored were more intensive in manual skills than in communication and cognitive skills.

In sum, offshoring reduces total employment and native employment, whereas immigrant workers complement native workers, and immigration generates positive productivity effect on total employment. Both native and immigrant workers upgrade their skills when faced with increased offshoring, and their average skill complexity also becomes more similar. Also, evidence from data description shows immigrant skill complexity is below native complexity. Based on these observations, I conclude that of native, immigrant and offshore workers, the average skill complexity is highest for natives, and lowest for offshore workers. And the fact that immigration expands native employment is due to the positive productivity effect generated by the influx of high-skilled immigrants. Although the complexity of tasks that are offshored is not observed, substitution between native and offshore workers may result from the competition between native and offshore workers performing less complex tasks, as those also appear more tradable.

3.5 Conclusion

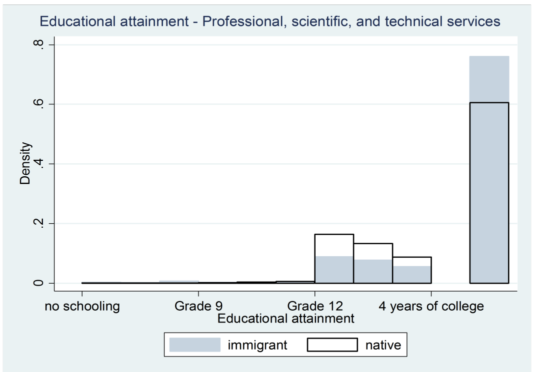
I employ data on US employment and offshoring in service sectors to measure the substitution effect and productivity effect of immigration and offshoring in this paper. Under the unified framework developed by Ottaviano, Peri, and Wright (2013), which allows joint analysis of the impact of offshoring and immigration, I analyze the relative skills of natives, immigrants and offshore workers through data on skill indices and the pattern of substitutability or complementarity between these three groups of workers.

Since the data on skill complexity of offshore workers are unobserved, I first look at the relative skills of natives and immigrants using the O*NET abilities data. Similar to the skill distribution in manufacturing industries, immigrants perform skills at lower complexity than natives. In addition, I find evidence of complementarity of immigrants and substitutability of offshore workers to natives from descriptive statistics. Next, I construct instruments for the cost of immigration and offshoring, and carry out regression analyses on the substitution and productivity effect. Regression estimates coincide with descriptive statistics in terms of substitution from offshore workers, and complementarity from immigrants. Furthermore, I offer an explanation that is consistent with the results from regressions, which imply that offshore workers are located at the lower end of the skill distribution. Based on the results, I speculate that the influx of immigrants are mostly composed of high-skilled ones, while the tasks offshored are of lower complexity.

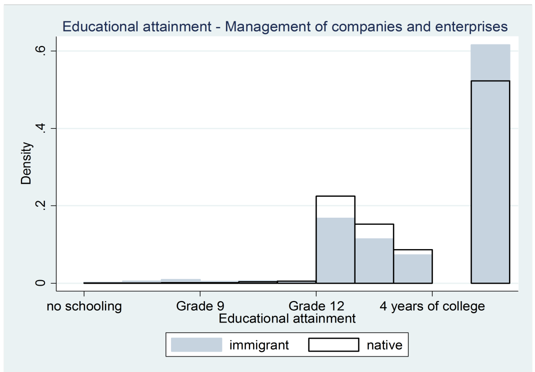
Although this paper provides various evidence on the substitution and productivity effects of offshoring and immigration, validity of the results may be undermined by the lack of variation in the data. More specifically, since the offshoring data at more disaggregated levels are unavailable, the estimates from 2SLS and direct OLS regressions may appear insignificant due to small sample size. Moreover, the service import variable might be endogenous. Thus we may need to consider other instruments for service offshoring. Future work may include construction of a theoretical model which adapts the framework in Ottaviano, Peri, and Wright (2013) to one with a different skill distribution. In addition, the results in this paper will be improved if data on less aggregated industry levels becomes available.

3.6 Figures and Tables

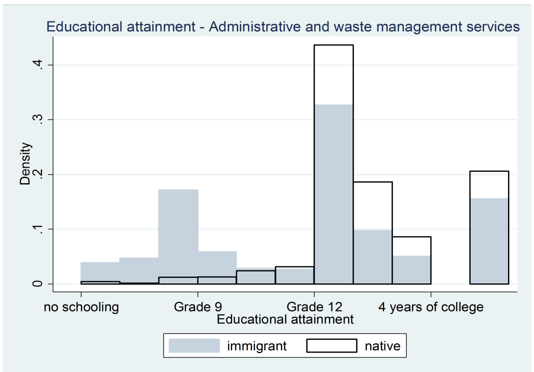
FIGURE 3.1: Immigrant employment shares and skill indices in 1,665 occupations, 2000-2009



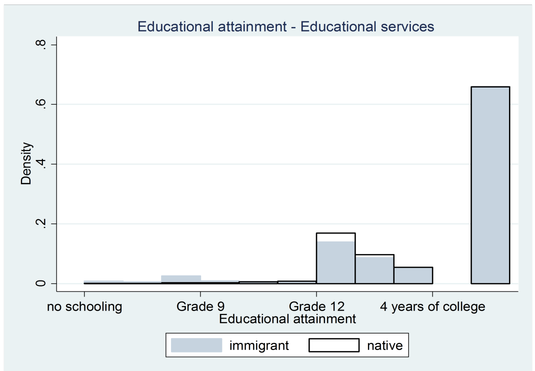
(A) Professional, Scientific and Technical Services



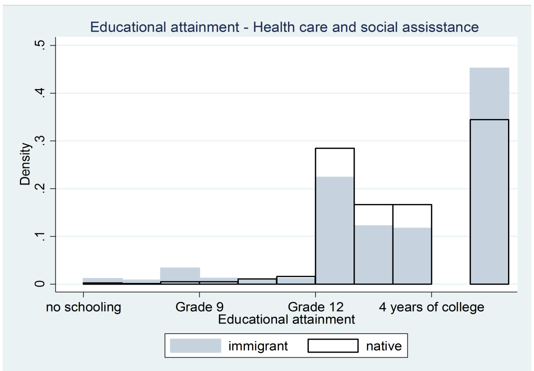
(B) Management of Companies and Enterprises



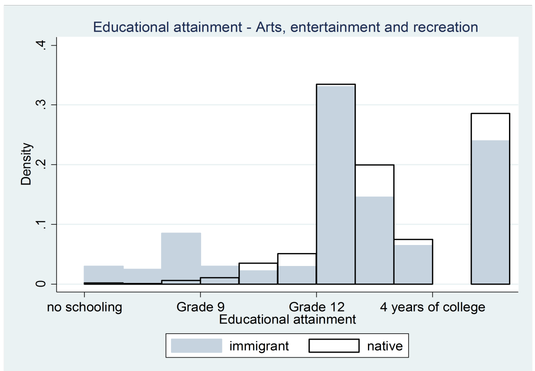
(C) Admin. and Waste Management Services



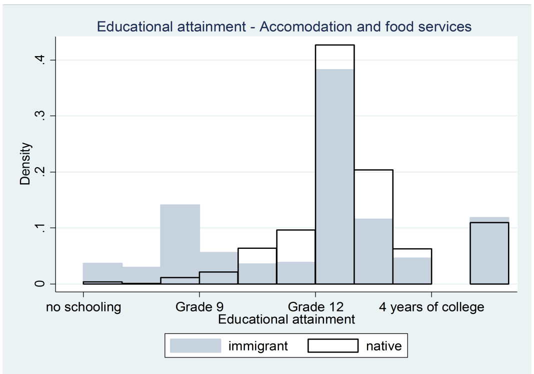
(D) Educational Services



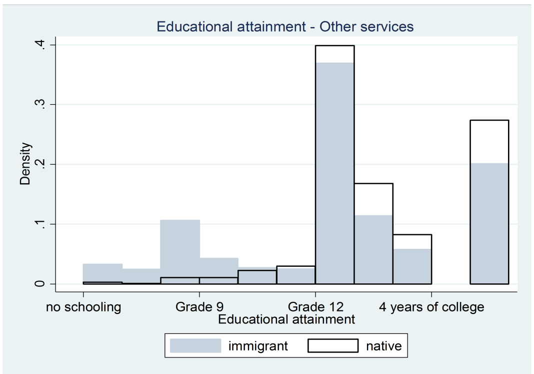
(E) Health Care and Social Assistance



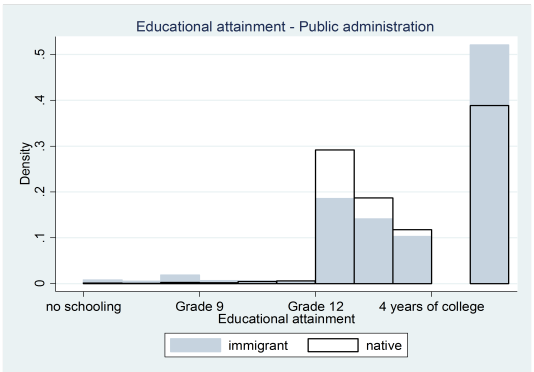
(F) Arts, Entertainment and Recreation



(G) Accomodation and Food Services

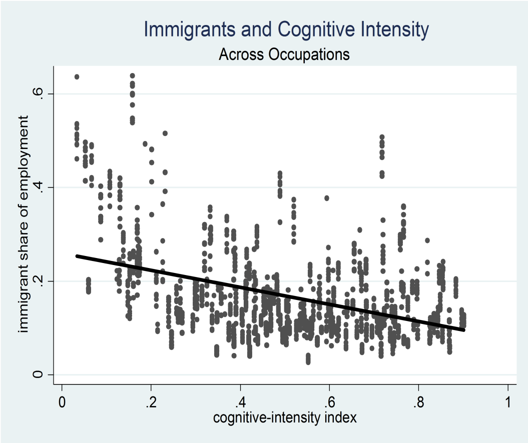


(H) Other Services

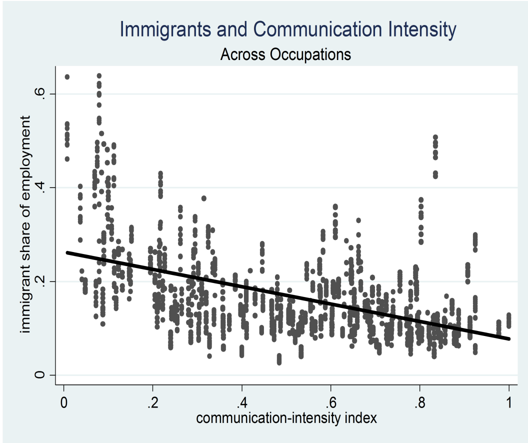


(I) Public Administration

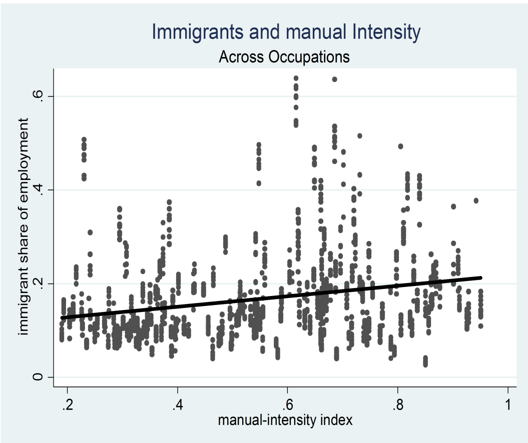
FIGURE 3.2: Immigrant employment shares and skill indices in 1,665 occupations, 2000-2009



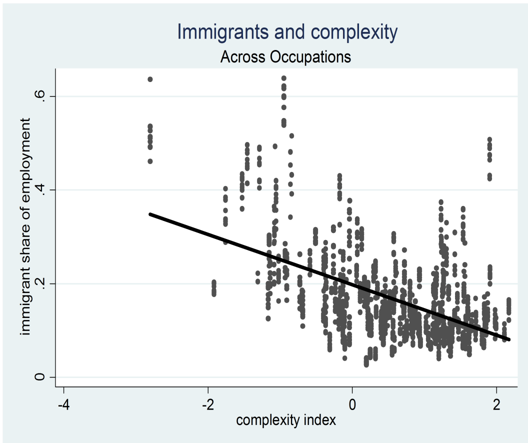
(A) Immigrant share and cognitive intensity; Slope=-0.18, Standard error=0.01



(B) Immigrant share and communication intensity; Slope=-0.19, Standard error=0.01

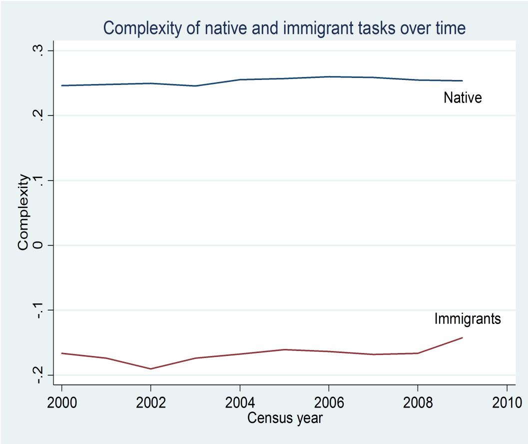


(C) Immigrant share and manual intensity; Slope=0.11, Standard error=0.01

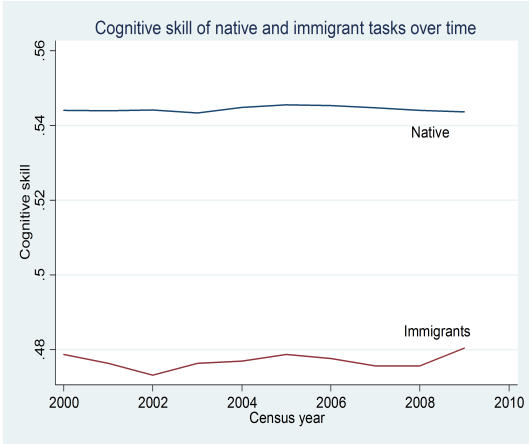


(D) Immigrant share and complexity index; Slope=-0.05, Standard error=0.003

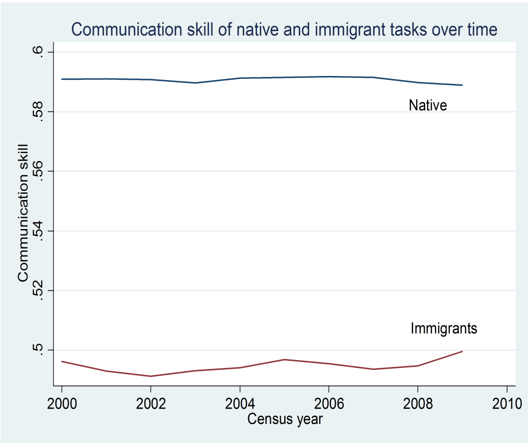
FIGURE 3.3: Evolution of skill indices for natives and immigrants, 2000-2009



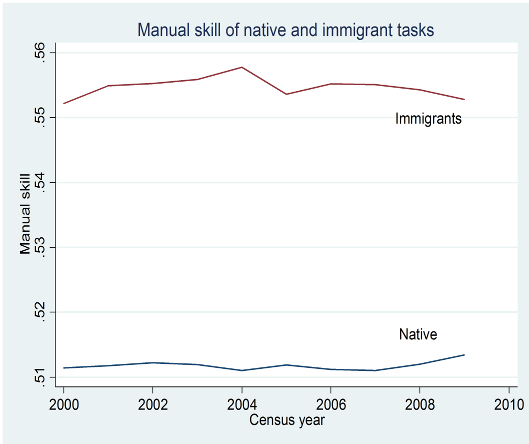
(A) Average complexity index over time



(B) Average cognitive intensity over time

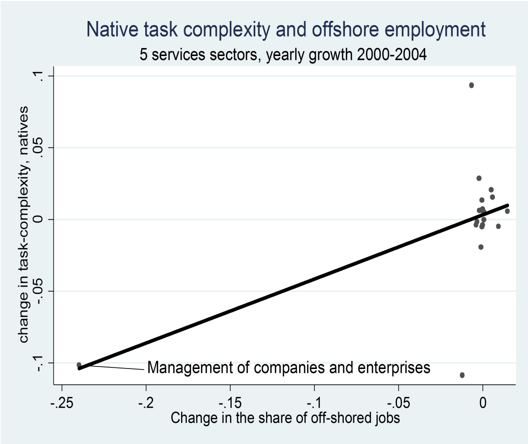


(C) Average communication intensity over time

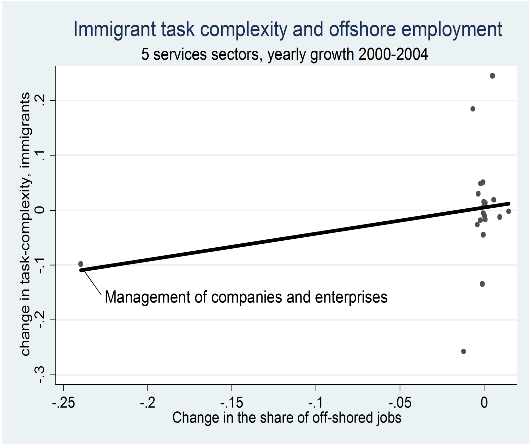


(D) Average manual intensity over time

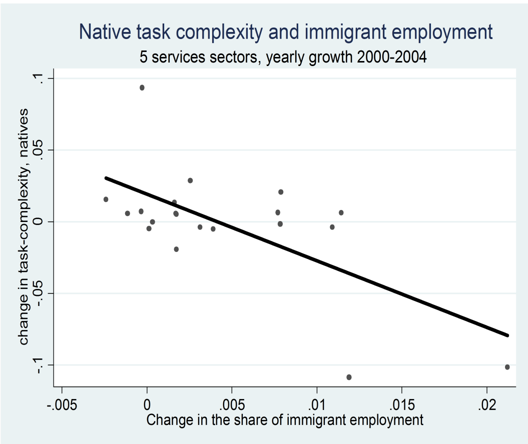
FIGURE 3.4: Task complexity and employment for native, immigrant and offshore workers



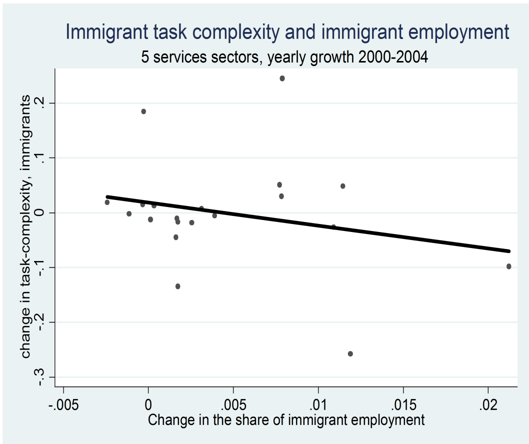
(A) Native complexity and offshore employment; Slope=0.45, Standard error=0.02



(B) Immigrant complexity and offshore employment; Slope=0.48, Standard error=0.1

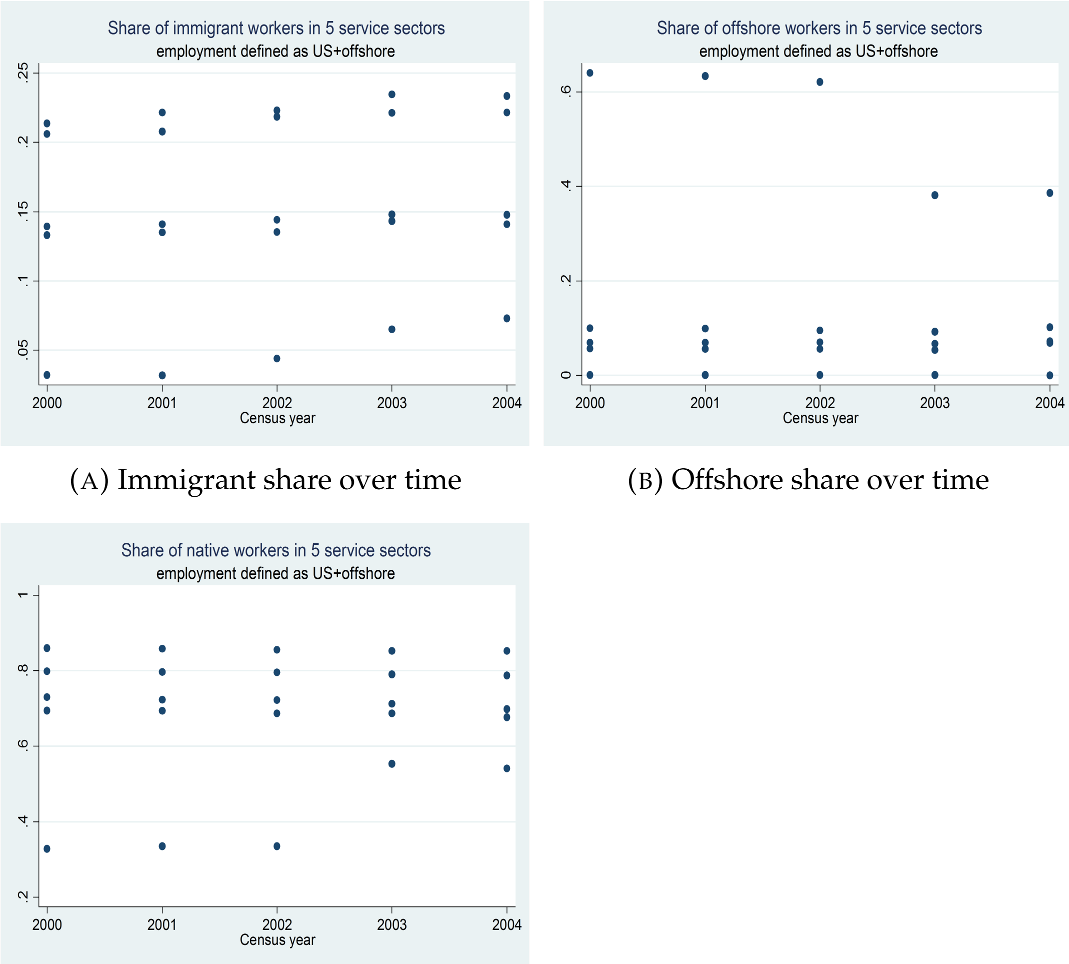


(C) Native complexity and immigrant employment; Slope=-4.65, Standard error=1.37



(D) Immigrant complexity and immigrant employment; Slope=-4.19, Standard error=3.57

FIGURE 3.5: Evolution of employment shares for native, immigrant and offshore workers in 5 service sectors

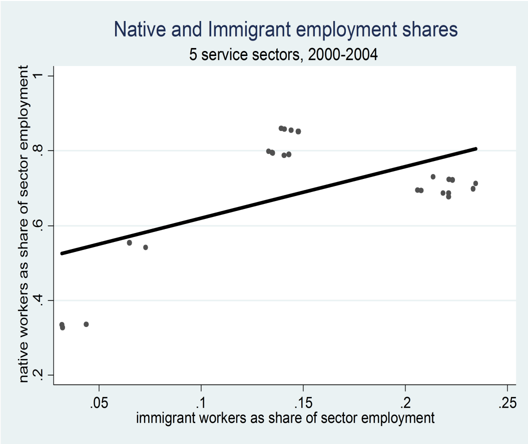


(A) Immigrant share over time

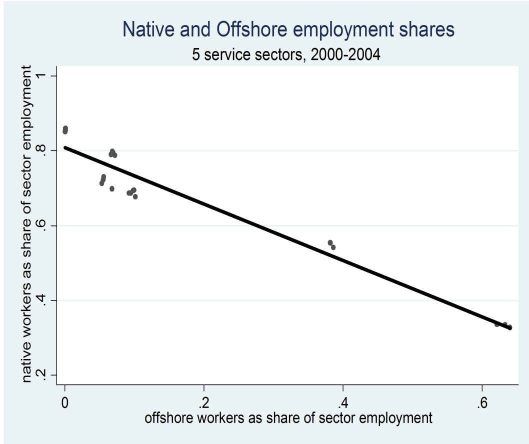
(B) Offshore share over time

(C) Native share over time

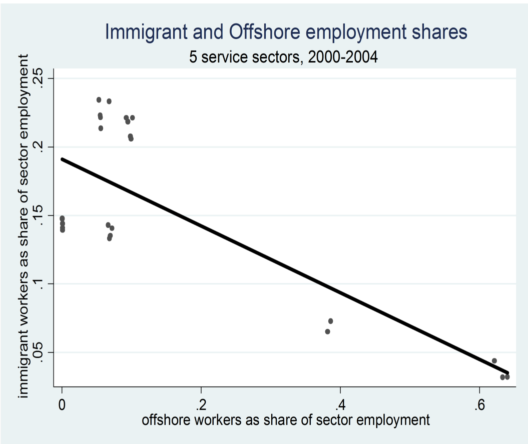
FIGURE 3.6: Correlation between employment shares of native, immigrant and offshore workers



(A) Native and immigrant employment shares; Slope=1.38, Standard error=0.46

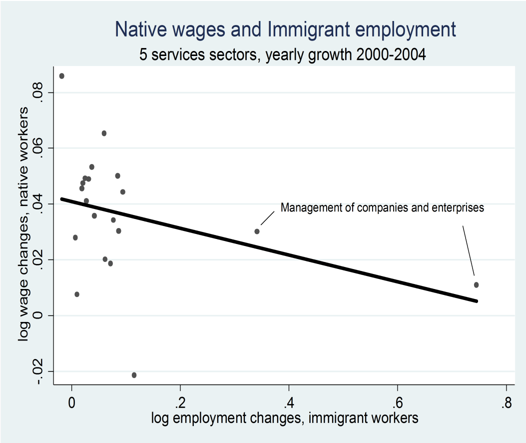


(B) Native and offshore employment shares; Slope=-0.76, Standard error=0.02

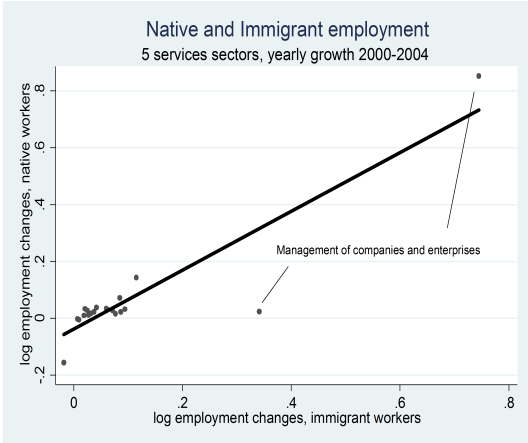


(C) Immigrant and offshore employment shares; Slope=-0.24, Standard error=0.02

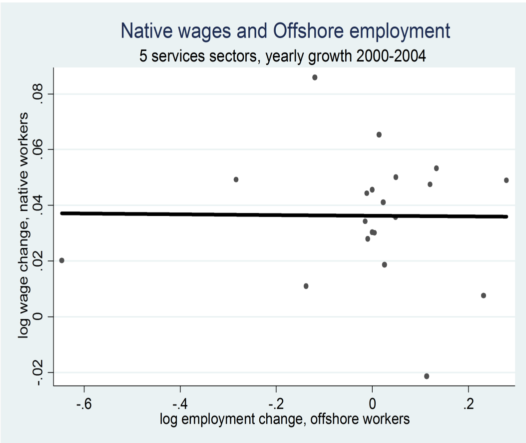
FIGURE 3.7: Correlations between native employment and wage and immigrant and offshore employment



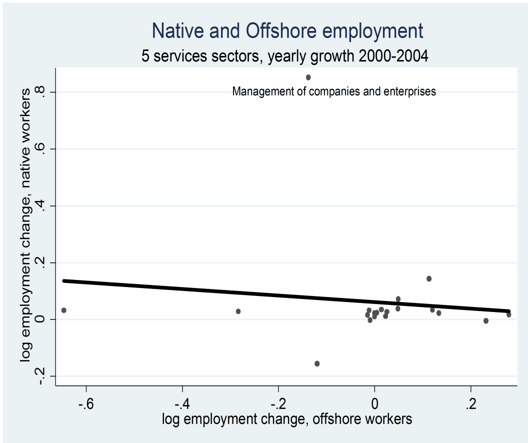
(A) Native wage and immigrant employment; Slope=-0.04, Standard error=0.01



(B) Native and immigrant employment; Slope=1.03, Standard error=0.2



(C) Native wage and offshore employment; Slope=-0.001, Standard error=0.02



(D) Native and offshore employment; Slope=-0.11, Standard error=0.218

TABLE 3.1: Effects of Immigration and Offshoring on Employment Shares

Specifications	Native Share of Employment				Immigrant Share of Employment		Offshore Share of Employment	
	IV	IV	OLS	OLS	IV	OLS	IV	OLS
IMM Share		20.04					-21	
		(10.5)					(10.5)	
OFF Share	-0.85***				-0.15***			
	(0.04)				(0.04)			
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
First Stage	OFF Share	IMM Share			OFF Share		IMM Share	
Imputed IMM share		-0.28		-5.56			-0.28	5.83
		(0.21)		(2.89)			(0.21)	(3.32)
IMP Service	0.62		-0.53		0.62	-0.09		
	(0.37)		(0.65)		(0.37)	(0.07)		
Observations	25	25	25	25	25	25	25	25
Wald F-stat of first stage	14.75	533.46	N/A	N/A	14.75	N/A	533.76	N/A

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Estimates are reported for OLS and IV estimations of 25 service sectors. Data on immigrants come from the Census and the American Community Survey, and data on employment come from the Bureau of Economic Analysis.

TABLE 3.2: Effects of Immigration and Offshoring on Employment Levels

Specifications	Log Native Employment				Log Immigrant Employment		Log Offshore Employment	
	IV	IV	OLS	OLS	IV	OLS	IV	OLS
Log IMM		0.98***					-0.28	
		(0.09)					(0.29)	
Log OFF	-1.76				-2.18			
	(1.58)				(1.88)			
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
First Stage	Log OFF	Log IMM			Log OFF		Log IMM	
Imputed IMM share		-20.72*		-20.25			-20.72*	5.74
		(6.09)		(10.71)			(6.09)	(5.73)
IMP Service	1.04		-1.84		1.04	-2.28		
	(0.82)		(2.18)		(0.82)	(2.59)		
Observations	25	25	25	25	25	25	25	25
Wald F-stat of first stage	1.65	11.57	N/A	N/A	1.65	N/A	11.57	N/A

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Estimates are reported for OLS and IV estimations of 25 service sectors. Data on immigrants come from the Census and the American Community Survey, and data on employment come from the Bureau of Economic Analysis.

TABLE 3.3: The Productivity Effects of Immigration and Offshoring on Total Employment

Specifications	Log Total Employment			
	IV	IV	OLS	OLS
IMM Share		28.2**		
		(16.06)		
OFF Share	-0.99			
	(0.36)			
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
First Stage	OFF Share	IMM Share		
Imputed IMM Share		-0.28		-7.82
		(0.21)		(3.96)
IMP Service	0.62**		-0.62	
	(0.37)		(0.69)	
Observations	25	25	25	25

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Estimates are reported for OLS and IV estimations of 25 service sectors. Data on immigrants come from the Census and the American Community Survey, and data on employment come from the Bureau of Economic Analysis.

TABLE 3.4: The Effects of Immigration and Offshoring on Task Complexity

Specifications	Complexity, native	Cognitive, native	Communication, native	Manual, native	Complexity, foreign born	Diff in Complexity, native vs foreign born
Panel A: 2SLS Estimates						
IMM Share	0.38	0.4	0.13	-0.03		
	(0.55)	(0.3)	(0.3)	(0.09)		
OFF Share	2.09**	0.07	0.01	-0.32	7.26	-4.83
	(0.97)	(0.1)	(0.2)	(0.32)	(6.52)	(4.99)
Panel A: OLS Estimates						
Imputed IMM Share	0.14	0.02	0.001	-0.005		
	(0.49)	(0.09)	(0.12)	(0.07)		
IMP Service	0.14	0.03	0.01	-0.02	0.46**	-0.3**
	(0.15)	(0.02)	(0.03)	(0.02)	(0.28)	(0.18)
Observations	25	25	25	25	25	25

Notes: Significance levels at 1%, 5% and 10% are represented by ***, ** and * respectively. Estimates are reported for OLS and IV estimations of 25 service sectors. Data on immigrants come from the Census and the American Community Survey, and data on employment come from the Bureau of Economic Analysis.

TABLE 3.5: List of Service Sectors and NAICS codes

Service Sectors	NAICS	
Professional, scientific and technical services (54)	5411	Legal Services
	5412	Accounting, tax preparation, bookkeeping and payroll services
	5413	Architectural, engineering, and related services
	5414	Specialized design services
	5415	Computer systems design and related services
	5416	Management, scientific and technical consulting services
	5417	Scientific research and development services
	5418	Advertising and related services
	54194	Veterinary services
	5419	Other professional, scientific and technical services
Management of companies and enterprises (55)	551	Management of companies and enterprises
Administrative and waste management services (56)	5613	Employment Services
	5614	Business support services
	5615	Travel arrangements and reservation services
	5616	Investigation and security services
	5617	Services to buildings and dwellings
	56173	Landscaping services
	5611	Other administrative, and other support services
	562	Waste management and remediation services
Educational Services (61)	6111	Elementary and secondary schools
	6112, 6113	Colleges, including junior colleges, and universities
	6114, 6115	Business, technical, and trade schools and training
	6116, 6117	Other schools, instruction and educational services
Health Care and Social Assistance (62)	6211	Offices of physicians
	6212	Offices of dentists
	62131	Office of chiropractors
	62132	Offices of optometrists
	6213	Offices of other health practitioners
	6214	Outpatient care centers
	6216	Home health care services
	6215, 6219	Other health care services
	622	Hospitals
	6231	Nursing care facilities
	6232	Residential care facilities, without nursing
	6241	Individual and family services
	6242	Community food and housing, and emergency services
	6243	Vocational rehabilitation services
	6244	Child day care services

Service Sectors	NAICS	
Arts, entertainment and recreation (71)	711	Independent artists, performing arts, spectator sports and related industries
	712	Museums, art galleries, historical sites, and similar institutions
	71395	Bowling centers
	713	Other amusement, gambling, and recreation industries
Accommodation and food services (72)	7211	Traveler accommodation
	7212, 7213	Recreational vehicle parks and camps, and rooming and boarding houses
	722	Restaurants and other food services
	7224	Drinking places, alcohol beverages
Other services (81)	8111	Automotive repair and maintenance
	811192	Car washes
	8112	Electronic and precision equipment repair and maintenance
	8113	Commercial and industrial machinery and equipment repair and maintenance
	8114	Personal and household goods repair and maintenance
	81143	Footwear and leather goods repair
	812111	Barber shops
	812112	Beauty salons
	812, 813	Nail salons and other personal care services
	8123	Drycleaning and laundry services
	8122	Funeral homes, cemeteries and crematories
	8129	Other personal services
	8131	Religious organizations
	8132	Civic, social, advocacy organizations and grantmaking and giving services
	81393	Labor unions
	8139	Business, professional, political and similar organizations
	814	Private households
Public administration (82)	92111	Executive offices and legislative bodies
	92113	Public finance activities
	92119	Other general government and support
	922	Justice, public order, and safety activities
	923	Administration of human resource programs
	924, 925	Administration of environmental quality and housing programs
	926, 927	Administration of economic programs and space research

Bibliography

Altonji, Joseph and David Card (1991). "The effects of immigration on the labor market outcomes of less-skilled natives". In: *Immigration, trade and the labor market*. University of Chicago Press, pp. 201–234.

Anderson, James (1979). "A Theoretical Foundation for the Gravity Equation". In: *American Economic Review* 69 (1), pp. 106–116.

Anderson, James and Eric Van Wincoop (2003). "Gravity with Gravitas: A Solution to the Border Puzzle". In: *American Economic Review* 93 (1), pp. 170–192.

Autor, David H., David Dorn, and Gordon H. Hanson (2013). "The China syndrome: Local Labor Market Effects of Import Competition in the United States". In: *American Economic Review* 103 (6), pp. 2121–68.

Blanchard, Olivier Jean and Lawrence F. Katz (1992). "Regional Evolutions". In: *Brookings Papers on Economic Activity* (1), pp. 1–61.

Borjas, George J (2003). "The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market." In: *National Bureau of Economic Research* w9755.

Borjas, George J. and Lawrence F. Katz (2007). "The evolution of the Mexican-born workforce in the United States". In: *Mexican immigration to the United States*. University of Chicago Press.

Bryan, Gharad and Melanie Morten (2017). "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia". In: *NBER Working Papers 23540*.

Card, David (2001). "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration". In: *Journal of Labor Economics* 19 (1), pp. 22–64.

Crino, Rosario (2010). "Employment effects of service offshoring: Evidence from matched firms". In: *Economics Letters* 107.2, pp. 253–256.

Dix-Caneiro, Rafael and Brian Kovak (2017). "Trade Liberalization and Regional Dynamics". In: *American Economic Review*.

Eaton, Jonathan and Samuel Kortum (2002). "Technology, Geography, and Trade". In: *Econometrica* 70, pp. 1741–1779.

Ebenstein, Avraham et al. (2014). "Estimating the Impact of Trade and Offshoring on American Workers using the Current Population Surveys". In: *The Review of Economics and Statistics* 100 (3), pp. 581–595.

Glaeser, Edward L. and Joseph Gyourko (2005). "Urban Decline and Durable Housing". In: *Journal of Political Economy* 113 (2), pp. 345–75.

Greenland, Andrew, John Lopresti, and Peter McHenry (2017). "Import Competition and Internal Migration". In: *Accepted at Review of Economics and Statistics*.

Grossman, Gene M. and Esteban Rossi-Hansberg (2008). "Trading tasks: A Simple Theory of Offshoring". In: *American Economic Review* 98(5), pp. 1978–1997.

Hakobyan, Shushanik and John McLaren (2016). "Looking for Local Labor–Market Effects of NAFTA". In: *Review of Economics and Statistics* 98:4.

Harrison, Ann and Margaret McMillan (2011). "Offshoring jobs? Multinationals and US manufacturing employment". In: *Review of Economics and Statistics* 93.3, pp. 857–875.

Head, Keith and Thierry Mayer (2014). "Gravity Equations:Gravity Equations: Workhorse,Toolkit, and Cookbook". In: *Handbook of International Economics*. Elsevier.

Head, Keith and John Ries (2008). "FDI as an Outcome of the Market of Corporate Control: Theory and Evidence". In: *Journal of International Economics* 74 (1), pp. 2–20.

Hsieh, Chang-Tai et al. (2016). "The Allocation of Talent and US Economic Growth". In: *Working Paper*.

Hummels, David et al. (2014). "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data". In: *American Economic Review* 104(6), pp. 1597–1629.

Kaplan, Greg and Sam Schulhofer-Wohl (2017). "Understanding the Long-run Decline in Interstate Migration". In: *International Economic Review* 58 (1), pp. 57–94.

Kerr, William R. and William F. Lincoln (2010). "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention". In: *Journal of Labor Economics* 28(3), pp. 473–508.

Kovak, Brian (2013). "Regional Effects of Trade Reform: What is the Correct Measure of Liberalization?" In: *American Economic Review* 103 (5), pp. 1960–1976.

Kurzendoerfer, Amanda (2015). "Essays in International Economics". PhD thesis. University of Virginia. PhD thesis.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak (2011). "Internal Migration in the United States". In: *Journal of Economic Perspectives* 25 (3), pp. 173–96.

Olney, William W (2012). "Offshoring, immigration, and the native wage distribution". In: *Canadian Journal of Economics* 45.3, pp. 830–856.

Ottaviano, Gianmarco I. P., Giovanni Peri, and Greg C. Wright (2013). "Immigration, Offshoring, and American Jobs". In: *American Economic Review* 103(5), pp. 1925–59.

Partridge, Mark D. et al. (2012). “Dwindling U.S. Internal Migration: Evidence of Spatial Equilibrium or Structural Shifts in Local Labor Markets?” In: *Regional Science and Urban Economics* 42, pp. 375–388.

Peri, Giovanni, Kevin Shih, and Chad Sparber (2015). “STEM Workers, H-1B Visas, and Productivity in US Cities”. In: *Journal of Labor Economics* 33, S225–S255.

Peri, Giovanni and Chad Sparber (2009). “Task specialization, immigration, and wages”. In: *American Economic Journal: Applied Economics* 1.3, pp. 135–169.

Pierce, Justin R. and Peter K. Schott (2012). “The Surprisingly Swift Decline of U.S. Manufacturing Employment”. In: *NBER Working Papers 18655*.

Portes, Richard and Helene Rey (2005). “The Determinants of Cross-border Equity Flows”. In: *Journal of International Economics* 65.2, pp. 269–296.

Ramos, Raul and Jordi Surinach (2016). “A Gravity Model of Migration between the ENC and the EU”. In: *Journal of Economic and Social Geography* 108.1, pp. 21–35.

Silva J., Santos and Silvana Tenreyro (2006). “The Log of Gravity”. In: *Review of Economics and Statistics* 88 (4), pp. 641–658.

Tinbergen, Jan (1962). “Shaping the World Economy: Suggestions for an International Economic Policy.” In: *Twentieth Century Fund, New York*.

Topel, Robert H (1986). “Local Labor Markets”. In: *Journal of Political Economy* 94 (3), S111–43.

Wright, Greg C (2014). “Revisiting the employment impact of offshoring”. In: *European Economic Review* 66, pp. 63–83.

Yuskavage, Robert E., Erich H. Strassner, and Gabriel W. Medeiros (2009). “Outsourcing and imported services in BEA’s industry accounts”. In: *International Trade in Services and Intangibles in the Era of Globalization*. University of Chicago Press.