ESSAYS IN LABOR ECONOMICS

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Keshar M. Ghimire
July 2016

Examining Committee Members:
Dr. Douglas Webber, Advisory Chair, Department of Economics
Dr. Johanna Catherine Maclean, Department of Economics
Dr. Michael Leeds, Department of Economics
Dr. Matt Saboe, External Member, West Chester University
This dissertation, in the standard three-essay format, studies three distinct but closely related aspects of the United States labor markets. Chapter 1 attempts to identify the main drivers of potential migration to the United States by using administrative data from the United States Diversity Visa Lottery. Estimating fixed effects panel data models that control for time-invariant unobserved heterogeneity in source-country level determinants of potential migration, I find that income levels in source countries and educational attainment of the source-country population play important role in determining migration intentions. Specifically, a one percent increase in per capita Gross Domestic Product of a source country decreases the potential migration rate from that country to the US by 1.36%. Similarly, a one percent increase in the educational attainment of source population (measured as the percentage of population with at least secondary education) decreases potential migration rate by 1.16%. The results obtained in this chapter improve our understanding of the composition of US labor markets by identifying the most important socio-economic variables that drive migration to the US.

Chapter 2 estimates the causal impact of a change in supply of immigrant entrepreneurs on entrepreneurial propensities of natives. I draw data from the Annual Social and Economic Supplement of the Current Population Survey and use within-state variation in supply of immigrant entrepreneurs for identification. To address concerns of endogeneity in the supply of immigrant entrepreneurs, I take advantage
of a quasi-experiment provided by the State Children’s Health Insurance Program. I find that, on average, immigrants self-employed in unincorporated businesses have no discernible impact on self-employment propensities of natives. However, immigrants self-employed in incorporated businesses crowd in natives into incorporated self-employment. Specifically, a 1% increase in incorporated immigrant entrepreneurs increases the supply of incorporated native entrepreneurs by 0.11%. Furthermore, various sub-sample analyses demonstrate substantial heterogeneity in the impact of immigrant entrepreneurs on entrepreneurial propensities of natives. The results obtained in this chapter have important implications for policies related to immigration and entrepreneurship development.

Finally, Chapter 3 exploits the State Children’s Health Insurance Program to investigate the impact of publicly funded health insurance coverage for children on labor supply of adults. Using data from the Annual Social and Economic Supplement of the Current Population Survey and triple difference identification strategy, the analysis demonstrates that public health insurance for children decreases labor supply of women, both at the extensive and the intensive margin, but increases that of men at the extensive margin. The estimates obtained in this chapter highlight the labor supply distortions associated with welfare benefits.
ACKNOWLEDGEMENTS

I am grateful to my committee for believing in me and guiding me through the graduate school. Thanks to Dr. Webber for taking me under his wings from early on in the program, to Dr. Maclean for always being available to answer questions and provide feedback on research, and to Dr. Leeds for guidance and support on everything including funding, research, teaching, and accommodation. Thanks to Dr. Matt Saboe for kindly accepting to be the external reader on a short notice.

I am thankful to all my friends, Temple economics faculty, referees, discussants, and conference participants whose thoughtful comments helped shape my research.

I am indebted to my parents for their investment in my early education. I am thankful to my brother for taking on family errands in my absence. Finally, I am grateful to my wife for her patience and support throughout my graduate studies.
# TABLE OF CONTENTS

ABSTRACT iii  
ACKNOWLEDGEMENTS v  
LIST OF TABLES x  
LIST OF FIGURES xi  

1 DRIVERS OF POTENTIAL MIGRATION: EVIDENCE FROM THE UNITED STATES DIVERSITY VISA LOTTERY 1  
1.1 Introduction ................................................. 1  
1.2 Literature Review ........................................... 5  
1.3 Conceptual Framework ....................................... 9  
1.4 Data .......................................................... 11  
1.4.1 The Diversity Visa Lottery ................................. 12  
1.5 Estimation and Results ....................................... 18  
1.6 Alternative Specifications ................................... 22  
1.7 Conclusion .................................................. 25  

2 SUPPLY OF IMMIGRANT ENTREPRENEURS AND NATIVE ENTREPRENEURSHIP 26  
2.1 Introduction .................................................. 26  
2.2 Related Literature ........................................... 30
3.6 Conclusion .................................................... 80

References ................................................................ 82

Appendices ................................................................ 94

A ................................................................................. 95

A.1 Generous SCHIP, Health Insurance Coverage of Native Children and Eligibility Guidelines ............................................. 95

A.2 Industry Classification in CEPR Extracts ......................... 96
LIST OF TABLES

1.1 Summary Statistics ........................................... 12
1.2 Drivers of Potential Migration to the US: Results from Pooled Regression 21
1.3 Drivers of Potential Migration to the US: Results from Fixed Effects Model ........................................... 21
1.4 Results from Dynamic Model: GMM One-step Estimates ........... 24
1.5 Results from Dynamic Model: GMM Two-step Estimates ........... 24
2.1 Summary Statistics: Self-employed Natives ....................... 45
2.2 Summary statistics: Self-employed Immigrants ..................... 46
2.3 Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives: Linear Probability Model .................. 49
2.4 First-stage Regressions: Association between Generous SCHIP and Supply of Immigrant Entrepreneurs ............................ 51
2.5 Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives: Two Stage Least Squares Model ........... 55
2.6 Heterogeneity in the Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives across Skill Levels: TSLS Estimates .................................................. 56
2.7 Heterogeneity in the Impact of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives across Industries: Two Stage Least Squares Model 59
2.8 Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Native Men and Women ................. 62
LIST OF FIGURES

1.1 Total Annual Diversity Visa Applications and Visa Awards . . . . . . 16
1.2 Annual Diversity Visa Applications from Selected Countries . . . . 17

2.1 Supply of Immigrant Entrepreneurs and Native Self-employment Rate 47
2.2 Distribution of Immigrant Entrepreneurs across Industries . . . . . 58

3.1 Pre and Post SCHIP Labor Supply . . . . . . . . . . . . . . . . . . . . 75

A.1 State Eligibility Guidelines for Children’s Medicaid/SCHIP . . . . 97
CHAPTER 1

DRIVERS OF POTENTIAL MIGRATION: EVIDENCE FROM THE UNITED STATES DIVERSITY VISA LOTTERY

1.1 Introduction

Global stock of international migrants has consistently grown during the last two decades. According to United Nation’s Migration Report 2013, the number of international migrants worldwide rose by over 77 million (or by 50%) between 1990 and 2013. Perhaps not surprisingly, most of that growth occurred between 2000 and 2010. According to the same report, there were 232 million international migrants living in countries other than where they were born in the year 2013. Of the total 2013 stock, 59% lived in developed nations and the rest lived in developing regions of the world indicating that migrant inflow is not limited to developed nations. Furthermore, a worldwide survey by Gallup estimates that nearly 700 million adults from around the world want to migrate to another country permanently if they could, United States
being the number one desired destination. These figures not only show that the stock of international migrants is large, both in ‘the South’ and ‘the North’, but also that these numbers are likely to grow in future.¹

International migration has for long been a topic of interest to economists because it has substantial socio-economic consequences on migrant sending and migrant receiving countries. Previous research has shown that, in migrant sending countries, emigration influences growth in per capita income and real wages (Boyer, Hatton, & ORourke, 1994), wage distribution across age groups (Elsner, 2013), inequality (Docquier, Ozden, & Peri, 2014), and democracy (Docquier, Lodigiani, Rapoport, & Schiff, 2015) among other things. Similarly, in migrant receiving countries, immigration has been shown to influence a wide range of socio-economic characteristics including real wages, unemployment rate, economic growth, internal migration, government expenditure, and entrepreneurship (Docquier, Ozden, & Peri, 2014; Dustmann, Fabbri, & Preston, 2005; Borjas, 2006; Rowthorn, 2008).

Because migration phenomena, both domestic and international, have important economic consequences, a large portion of economic research on migration has focused on what drives people to migrate, often across vastly different cultures and languages. This paper contributes to the existing literature on determinants of international migration by identifying drivers of willingness to migrate based on administrative data from the United States Diversity Visa Lottery.

A number of theoretical perspectives such as those embodied in Neo-classical Macroeconomic Theory, Neo-classical Microeconomic Theory, New Economics of migration, and Social Capital Theory - to be discussed below - have been put forward to explain reasons behind movement of people within and across borders. There is also a proliferation of empirical studies attempting to identify drivers of interna-

¹It is common in migration literature to refer to developed nations as ‘the North’ and the less developed nations as ‘the South.’ Migration scholars are often concerned about distinction between South-North and South-South migrations.
tional migration. Although empirical approaches have been varied mainly due to data limitations, the standard approach in identifying the determinants of international migration empirically has been to estimate gravity models (of the type popular in international trade literature) that specify bilateral migrant flows as functions of a number of socioeconomic characteristics of sending and receiving countries. The intuition behind the standard approach is that there is a a set of ‘push factors’ that encourages individuals to leave their home countries and a set of ‘pull factors’ that attracts these individuals to destination countries. The push and pull factors with statistically significant coefficients in regressions are then interpreted as being the main drivers of international migration.

Keeping aside the usual econometric concerns related to data quality and endogeneity of regressors, an important limitation of the standard approach is a rather unrealistic assumption that individuals who want to migrate, conditional on the information they possess about their current location and potential destinations, can actually migrate. Therefore, recently there has been interest in separating the concept of migration aspirations from actual migration capabilities. Carling (2002), in his seminal paper on migration aspirations, emphasized the need to address migration intentions and actual migration separately.

The notion of distinct existence of migration aspiration prior to actual migration stems from the realization that not every individual who wants to migrate can migrate. Constraints like restrictive immigration policies of most receiving countries and high visa costs of western countries prevent a large number of aspiring migrants from realizing their destination preferences (Hatton & Williamson, 2002; Castles & Miller, 2003; Ratha & Shaw, 2007). The structural barriers such as border restrictions and wealth and/or financial constraints significantly dampen the volume of international migration leaving a pool of aspiring migrants stranded in their home countries against their will to stay (Paul, 2011). These observations have led migration scholars to por-
tray international migration as a two step process where agents first form willingness to migrate based on some utility maximization principle and then actually migrate only if they are able to overcome various constraints they face (Docquier, Peri, & Ruyssen, 2014).

Indeed, if cross border migration does involve two distinct steps as recognized in literature, extra insight can be gained by analyzing migration in two steps (Carling, 2002). For example, knowing the determinants of willingness to migrate can help forecast actual migration under different policy regimes such as open-borders versus semi-permeable borders. To that end, Docquier, Peri, and Ruyssen (2014) make a significant contribution to literature by analyzing potential and actual migration rates separately. The authors use data from a World Gallup surveys to find that stock of co-nationals and earnings at the destination are the most important drivers of potential migration. Moreover, analysing actual migrant flows while controlling for pool of potential migrants, the authors conclude that economic growth at destination and educational attainment of potential migrants are the major factors contributing to actual migration of the potential migrants. While a pioneer study to quantify the two steps of international migration, there are at least two important limitations of Docquier, Peri, and Ruyssen (2014). First, the study uses survey data whose questions are open to interpretation and second, the cross-sectional nature of the study leaves the authors unable to control for source-country level unobserved heterogeneity in determinants of potential and actual migration.

This paper extends the emerging literature on potential migration by identifying the drivers of aggregate potential migration rate to the United States using administrative data from US DV lottery. The US DV Lottery, also known as the Green card Lottery, is a diversity visa program which makes available up to 55,000 diver-

---

2Potential migration rate for any country is proportion of populations that is willing to migrate and the actual migration rate is the proportion of population that is able to migrate by overcoming the migration constraints.
sity visas for permanent residence in the US annually. Under this program, aspiring
migrants from around the world are allowed to file free applications for a chance to
win a permanent resident visa to the US.

One advantage of using these applications as a measure of willingness to migrate is
that the intention to migrate is unequivocally revealed. Another advantage is the use
of panel data in this context. Having panel data allows me to control for unobserved
country-level heterogeneity in the determinants of potential migration rate.

Estimating parsimonious fixed effects model, I find that per capita income at the
source countries and the educational attainment of the population at source are the
most important drivers of potential migration rate. More specifically, the results
generated in fixed effects models suggest that a 1% increase in per capita income
in source countries decreases the potential migration rate by 1.36% and that a 1%
increase in the percentage of population with at least secondary education decreases
the potential migration rate by 1.16%. The results obtained in this paper are robust
to dynamic specifications that allow for short run adjustments in response to deviation
from a long run equilibrium in potential migration rate.

The rest of the paper is organized as follows: next section provides a brief review of
the literature on determinants of international migration, Section 2.3 provides a sim-
ple conceptual framework for determining drivers of potential migration, Section 2.4
discusses data, Section 2.5 presents estimation and results, and Section 2.6 concludes.

1.2 Literature Review

Existing literature on determinants of international migration is massive and an ex-
haustive discussion of all important works on the topic is out of the scope of this
thesis. Following paragraphs outline only some of the influential works in the area.

A recorded theoretical analysis of migration is said to a have begun with Smith
where Smith observed that wage differences between rural and urban areas that are not accounted for by the differences in prices lead to migration of labor. After a long hiatus in economic analysis of determinants of migration, Hicks (1932) claimed that differences in wages were the main reason behind migration of individuals. However increasing wage disparities across labor markets thereafter showed clearly that things other than wages were also important for determining migration. As a result, most modern theories of migration build on the ideas of Smith and Hicks and theorize that a wide range of other social and economic issues, in addition to wages, have role in determining migrant flow.

Modern theories of international migration can be discussed under four headings. The Neo-classical Macroeconomic theory of migration posits that migration is merely a movement of labor due to geographical differences between demand and supply of labor (Bauer & Zimmermann, 1999). While the basic tenets of Macroeconomic theory of migration were discussed by Smith (1776), Ravenstein (1889) and Hicks (1963), Harris and Todaro (1970) enriched, formalized and popularized the Neo-classical Macroeconomic models.

The Neo-classical Microeconomic theory of migration is often credited to Sjaastad (1962) who presented migration decision as an individual human capital investment problem. According to this theory, individuals perform a cost benefit analysis of migration based on returns to their human capital in multiple labor markets. The core of the theory is that individuals rank multiple destinations on the basis of returns to their idiosyncratic human capital and move to the place where they expect highest returns to their skills. Some variations of neo-classical microeconomic theories go beyond assuming income differentials as the sole driver of migration decisions by including other amenities that might add to individuals utility functions.

New Economics of Migration Theory, put forth by Stark and Bloom (1985), claims that migration decisions are taken at household level rather than individual level.
Therefore, according to this theory, the main intention is not to maximize expected income of an individual but to diversify risk in revenues of a household. Operating in different labor markets reduces the impact of a particular income shock on the whole family.

The Social Capital Theory of migration posits that the network for family and friends present in destination has a major impact on the future migration flow. This has been discussed by E. S. Lee (1966) and Massey et al. (1993). According to this theory, the social capital at the destination, often in terms of network of families, friends, and co-nationals, helps the potential migrants in pre-migration information acquisition, post migration accommodation and search for employment opportunities, etc.

Empirical studies on determinants of international migration often take the approach of regressing bilateral migrations flows over a number of hypothesized determinants of migration. The goal is to identify the most important determinants from among the proposed set of determinants. In some sense the empirical studies intend to test the implications of the theories of migration discussed earlier.

Empirical studies have looked at determinants of internal as well as international migration. For example Borjas, Bronars, and Trejo (1992) use data from National Longitudinal Survey of Youth (NLSY) to study determinants of migration within the United States and find that interstate differences between returns to skills are the major determinants of migration. Adams Jr (1993) uses data from rural Egypt to ‘identify economic and demographic determinants of international migration’ and finds that the relationship between income and migration is an inverted U-shaped curve, and that poverty and landlessness are the major push factors sending rural Egyptians out of the country. More recently Mayda (2010) uses panel data of migration inflows into Organisation for Economic Co-operation and Development (OECD) countries to examine the ‘impact of geographical, cultural, and demographic factors
as well as the role played by changes in destination countries’ migration policies.

The author finds that ‘pull factors’ such as income levels at the destination are much more important in determining migration flows than the ‘push factors’ such as bad economic conditions in source countries. Similarly Ortega and Peri (2013) use panel data of bilateral flows from 120 sending countries to 15 OECD receiving countries. By employing an empirical model that accounts for availability of multiple destinations, the trio find that migration flows are most responsive to per capita income at destination.

Most relevant for this study is the recent paper by Docquier, Peri, and Ruyssen (2014) which conceptualizes international migration as a two step process where agents form willingness to migrate based on a pre-migration cost-benefit analysis and then actually migrate only if the migration opportunities such as visas become available. The authors use answers from the Gallup Poll surveys - that ask adults if they would migrate if given a chance - to measure aggregate potential migration rate from countries around the world. The authors, then, regress the aggregate potential migration rates on a number of source and destination country characteristics to find that the income level and employment probability at destination combined with network of co-nationals living abroad are major determinants of potential migration rates. Moreover, by estimating a separate set of regressions for actual migration rate, the authors find that the educational achievement of the potential migrants and growth prospects at the destination rate are major drivers of actual migration rate. Although important as a pioneer study in this tradition of analysing migration as a two step process, the Docquier et. al. study, as stated earlier, has two important limitations. First the survey data is vulnerable to problems related to respondents’ interpretation of the questions asked. Second, the use of cross-sectional data makes authors unable to control for unobserved heterogeneity across countries. This paper extends the emerging literature on potential migration not only by providing evi-
dence based on administrative data but also by providing more reliable estimates of
the elasticities of potential migration derived from panel data models.

1.3 Conceptual Framework

The drivers of potential migration can be analysed under a conceptual framework
that is not much different from the traditional framework used in previous empirical
studies of drivers of international migration. For simplicity, let’s consider a two-
country setup discussed in Borjas (1987a). The residents of the source country have
some information regarding socioeconomic characteristics of their home country and
potential destination country. The utility maximization principle dictates that in-
dividuals in source country are willing to migrate if they expect a higher utility in
a destination country after accounting for the cost of migration. Therefore an indi-
vidual’s willingness to migrate is a function the socioeconomic characteristics of the
home country, those of the destination country, and the cost of move.

Following Borjas (1987a), let’s assume, for a moment, that income is the sole
argument in the utility function. Then, individuals in a source country face the
earnings distribution:

\[ \log w_0 = \mu_0 + \nu_0 \]  \hspace{1cm} (1.1)

where \( w_0 \), the wage the worker earns in home country, depends on \( \mu_0 \), the mean
log wage in source country, and a normally distributed deviation term \( \nu_0 \). If the
whole population of the source country were to migrate to the host country, then the
migrants would face the earnings distribution:

\[ \log w_1 = \mu_1 + \nu_1 \]  \hspace{1cm} (1.2)
where $w_1$, $\mu_1$, and $\nu_1$ are the worker’s wage, mean log wage, and random deviation from mean log wage in destination country.

If migration decisions are mainly determined by comparing earnings opportunities across countries after accounting for migration costs, the probability of migration can be modelled first by defining an index function $I$ such that:

$$I = \log\left(\frac{w_1}{w_0} + C\right) \approx (\mu_1 - \mu_0 - \pi) + (\nu_1 - \nu_0) \quad (1.3)$$

where $C$ is the level of migration costs, and $\pi$ gives the time-equivalent measure of these costs. A person wants to emigrate if $I > 0$, and not if otherwise. The random variable $\pi$ is distributed:

$$\pi = \mu_{\pi} + \nu_{\pi} \quad (1.4)$$

where $\mu_{\pi}$ is the mean level of migration costs in the population and $\nu_{\pi}$ is a normally distributed random variable with mean zero and variance $\sigma_{\nu}^2$.

The probability that an individual wants to migrate to destination country can now be written as:

$$P(z) = Pr[\nu > -(\mu_1 - \mu_0 - \mu_{\pi})] = 1 - F(z) \quad (1.5)$$

where $\nu = \nu_1 - \nu_0 - \nu_{\pi}$, $z = -(\mu_1 - \mu_0 - \mu_{\pi})/\sigma_{\nu}$, and $F(z)$ is the standard normal distribution function (Borjas, 1987a).

A few testable implication of the outlined theory, as pointed out by Borjas (1987a), are $\partial P/\partial \mu_0 < 0$, $\partial P/\partial \mu_1 > 0$, and $\partial P/\partial \mu_{\pi} < 0$. In other words, the potential migration rate decreases when mean income in the source country rises, mean income in receiving country falls and when migration costs rise. These implications can be tested in a regression framework by approximating $F(.)$ with a conditional expectation function. The role of other potential determinants of willingness to migrate can be
tested by including the potential correlates as additional variables in the regression function for $P(.)$.

### 1.4 Data

This study uses publicly available data from multiple sources. The key variable to be explained in this study is potential migration rate. I define the potential migration rate for a country as the annual number of diversity visa applications filed from the country per 1000 adult population.

Data on total number of applications by country of chargeability for the diversity visa programs of the years 2007 to 2013 is available on the website of the state department of the United States. Data outside this period is not available publicly as of this draft. I retain only countries/territories which are listed on Penn World Table (PWT 7.1) to make sure I have data on relevant economic variables for countries included in the analysis.\(^3\) I also drop the countries which were eligible for diversity visa applications for less than 5 years during the given period to prevent a few outliers from driving the main results of the paper. That leaves me with qualified annual applications from 174 countries for further analysis. Since the applications for diversity visa are filed 2 years before the program-year, I have annual numbers filed in the years between 2005 to 2011.

The explanatory variables used in the study come from a variety of sources. Data on total adult population and 15-29 age group is obtained from United Nation’s ‘World Population prospects: The 2012 Revision.’ Purchasing Power Parity (PPP) converted per capita Gross Domestic Product (GDP), in 2005 international dollars, are obtained from Penn World Tables, data on political and civil liberties come from Freedom House, the internet users per 100 people in source country population from the World

\(^3\)See Heston, Summers, and Aten (2012) for details on PWT 7.1.
Bank, and educational attainment data from Barro and Lee (2013). One downside of putting together the final dataset from data obtained from multiple sources is that the number of complete observations available for analysis is reduced drastically. The summary statistics on the relevant variables, along with corresponding number of available records, are presented in Table 1.1. The following subsection will briefly introduce the United States Diversity Visa Lottery.

Table 1.1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential migration rate</td>
<td>1124</td>
<td>3.335</td>
<td>7.842</td>
</tr>
<tr>
<td>Source per capita GDP</td>
<td>1169</td>
<td>8353.42</td>
<td>9985.70</td>
</tr>
<tr>
<td>US per capita GDP</td>
<td>1336</td>
<td>45240.53</td>
<td>2721.048</td>
</tr>
<tr>
<td>Source political and civil liberties</td>
<td>1640</td>
<td>4.51</td>
<td>1.967</td>
</tr>
<tr>
<td>Immigrant stock</td>
<td>907</td>
<td>566.346</td>
<td>587.129</td>
</tr>
<tr>
<td>Pop. between 15 and 29</td>
<td>1336</td>
<td>113699.3</td>
<td>146396.10</td>
</tr>
<tr>
<td>Pop. w/ at least secondary education (%)</td>
<td>1640</td>
<td>28.772</td>
<td>15.61</td>
</tr>
<tr>
<td>Internet users at source (%)</td>
<td>1296</td>
<td>20.383</td>
<td>21.69</td>
</tr>
</tbody>
</table>

1.4.1 The Diversity Visa Lottery

United States Diversity Visa Lottery, also known as the Green card Lottery, is a diversity visa program which makes available up to 55,000 permanent resident diversity visas annually.\(^4\) The program is congressionally mandated by the section 203(c) of the Immigration and Nationality Act (INA) of 1990 (USCIS).

Background

The Diversity Visa program became the law in 1990 when diverse immigrant category was added to the Immigration and Nationality Act to attract more diverse

\(^4\)The Nicaraguan and Central American Relief Act (NCARA) passed by Congress in November 1997 stipulates that 5,000 of the 55,000 annually allocated diversity visas will be made available for use under the NCARA program.
migrants from parts of the world that were not heavily represented in the United States. The diversity visa lottery program in current form began in 1995 after three transitional years with temporary lotteries. From 1992 to 1994, the Department of State conducted lottery for 40,000 immigrant visas to be awarded to natives for countries which were adversely affected by the the 1965 amendments of Immigration and Nationality Act that ended the country quota system. 40% of the transitional visas each year were reserved for natives of Ireland alone (Wasem & Ester, 2004).

From fiscal year 1995 to 2004, the program was conducted in a ‘mail-in’ format. Applicants were required to mail-in the entries to a designated address and the entries had to arrive within a specified month-long period during the Fall of each year. For example, the DV-2000 registration mail-in period was held from noon on Thursday, October 1, 1998 through noon on Saturday, October 31, 1998. And the entries received before or after the specified registration dates regardless of when they are postmarked and entries sent to an address other than one of those indicated were considered void (U.S. Department of State, 1998).

No designated application forms were required but the entry had to be typed or clearly printed in the English Alphabet on a sheet of plain paper and had to include: i) Applicant’s full name, spouse’s name and dependents’ names. ii) Applicant’s mailing address and native country, iii) A recent (preferably less than 6 months old) photograph (1 1/2 inches by 1 1/2 inches) of the principal applicant with the applicant’s name printed across the back of the photograph, and iv) Applicant’s signature in the native language. Successful candidates chosen by a random lottery were notified by mail.

Since DV- 2005, however, the program has been accepting only electronic applications. The applications for DV 2005 were accepted during the 60-day period that ran from 12:00 AM November 1, 2003 to midnight, December 30 (U.S. Department of State, 2004). Beginning DV-2012 the application period was cut to a 30 day pe-
period, usually starting early October and ending in early November. For example, the DV-2012 application period ran from October 5, 2010, until noon, November 3, 2010.

Eligibility

Natives of countries with historically low rates of immigration to the United States are eligible to take part in the DV lottery program. Countries with historically low rate of immigration are those sending less than 50,000 immigrants (through family sponsored and employment based pathways) in the previous 5 years. This results in eligibility of particular countries change over years.

For DV-1995, natives of CHINA - mainland born and Taiwan born, India, South Korea, Philippines, Vietnam, UK (except Northern Ireland), Canada, Dominican Republic, El Salvador, Jamaica, and Mexico were not eligible to apply for Diversity Visa lottery (*U.S. Department of State*, 1994). The list of ineligible countries has very slowly evolved over the years, previously ineligible countries being eligible during certain years and vice versa. For example the following countries were not eligible to participate in DV-2013: Bangladesh, Brazil, Canada, China (mainland-born, excluding Hong Kong S.A.R., Macao S.A.R., and Taiwan), Colombia, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, India, Jamaica, Mexico, Pakistan, Peru, the Philippines, South Korea, United Kingdom (except Northern Ireland) and its dependent territories, and Vietnam. Similarly, for DV-2015, natives of the following countries were not eligible to apply, because more than 50,000 natives of these countries immigrated to the United States in the previous five years: Bangladesh, Brazil, Canada, China (mainland-born), Colombia, Dominican Republic, Ecuador, El Salvador, Haiti, India, Jamaica, Mexico, Nigeria, Pakistan, Peru, Philippines, South Korea, United Kingdom (except Northern Ireland) and its dependent territories, and Vietnam.

It is important to note, however, that if someone was not born in an eligible
country, then there still are two other ways to make own-self eligible. One can claim spouse’s country of birth if the latter is eligible or can claim country of birth of parents if one was born in an ineligible country while their parents were not born or not legal resident of the country.

Beside affiliation, DV applicants should also be able to provide proof of 12 years of education or two year experience in past 5 years in an occupation that requires at least two years of training or experience to perform at the time they apply for visa. This evidence is required only if chosen by the random lottery.

**Trends in Applications and Visas**

Millions of applicants have tried their luck since the Diversity Program in current form started in 1995. During the mail in period a millions of applications were disqualified because they arrived either too early or too late. For example 2.4 million entries filed for DV 1999 were disqualified. Similarly 2.5 millions entries filed for DV 2000, 2 million for DV 2001, 3 million for DV 2002, 2.5 million for DV 2003 and 2.9 million for DV 2004 were disqualified. It is reasonable to expect that the number of disqualified entries should have gone down when the Department of State transitioned to a fully electronic filing but the department doesn’t report the disqualified entries any more.

The Department of State reports qualified entries in its visa bulletin every year. Between DV-1995 and DV-2014, the number of qualified DV applications has gone up from 6.5 million to 9.4 million. The highest number (approximately 14.7 million) of qualified applications were received for DV 2012 and the least number (approximately 3.4 million) of qualified applications were received for DV-1999. Figure 1.1 shows the global trend in applications and the annual number of diversity visas awarded.

There is a considerable variation in the trends between countries that are eligible for diversity visa applications. For example, while Ghana shows a clearly increasing trend in total DV applications, Morocco shows a decreasing trend. At the same time
countries like Turkey and Kenya show a roughly constant number of diversity visa applications to DV 2007 through DV 2013. Figure 1.2 provides a snapshot of the variation in the trends between 9 different countries. This large variation in applications to diversity visa program can be exploited to deduce meaningful relationship between people’s willingness to migrate to the US and the possible drivers of such willingness.
Figure 1.2: Annual Diversity Visa Applications from Selected Countries
1.5 Estimation and Results

Following previous literature on determinants of international migration and the conceptual framework developed in section 3, I specify the following fixed effect model to identify the main drivers of potential migration:

\[
\text{pmr}_{j,t} = \alpha + \beta_1 \times \text{pcgdp}_{source,j,t} + \beta_2 \times \text{pcgdp}_{dest,t} + \beta_3 \times \text{average liberties}_{j,t} \\
+ \beta_4 \times \text{internet access}_{j,t} + \beta_5 \times \text{immigrant stock}_{j,t} + \beta_6 \times \text{age}_{15\text{to}29,j,t} \\
+ \beta_7 \times \text{percent seceduc}_{j,t} + \mu_j + u_{j,t} \quad (1.6)
\]

where \( \text{pmr}_{j,t} \) is potential migration rate for source country \( j \) at time \( t \) and is defined as the number of DV applicants from country \( j \) in the year \( t \) divided by the adult population (in thousands) of country \( j \) in the same year. \( \mu_j \) captures the source country fixed effects and \( u_{j,t} \) is measurement error. The \( \beta_s \) are the parameter of interest.

Only the controls found to be most relevant for international migration in previous studies are included. Specifically, I include income levels (in PPP adjusted per capita terms) in the source countries (\( \text{pcgdp}_{source} \)), income levels in the US (\( \text{pcgdp}_{dest} \)), average of political and civil liberties in source countries (\( \text{average liberties} \)), fraction of young (age group 15 to 29) population in source country (\( \text{age}_{15\text{to}29} \)), fraction of population with secondary education in source country (\( \text{percent seceduc} \)), and stock of immigrants in the US originating from the source country (\( \text{immigrant stock} \)). An additional control that is not commonly found in previous studies of determinants of international migration is the proportion of source country population with access to internet (\( \text{internet access} \)). This is an important control in this study because US DV essentially became Electronic Diversity Visa (EDV) in the years starting 2005,
and it is only reasonable to expect that general access to internet may have impact on the number of DV applications files. It is common in the migration literature to use distance (geographical, cultural, ethnic, etc.) between the nations as a proxy for cost of migration. Equation 1.6 captures such measures of cost, at least those that are fixed over time, through $\mu_j$.\footnote{This may seem like an unreasonable assumption in the sense that cultural distances between countries are dynamic variables and are likely to change over time. However, it is reasonable to assume that these are fairly static over the short duration of time considered in this study.}

A fixed effect model is intuitive in this setting to control for time invariant unobserved source country level heterogeneity in drivers of willingness to migrate. However, for the sake of comparison I also report estimates from a pooled regression in Table 1.2.

We obtain some common sense results from the pooled regressions. As implied by the neoclassical macro theories, the willingness to migrate is negatively related with average income at home. The coefficient on the income of the destination country has a positive sign, as predicted by the theory, but this coefficient is not statistically significant at conventional levels. The other statistically significant coefficients are on immigrant stock and educational attainment at the source. These coefficients suggest that willingness to migrate is higher if there is a greater network of fellow country men or women at the destination, as implied by social capital theory, and that better educated individuals are more willing to migrate, as implied by human capital investment theory. The proportion of youth in the population is barely significant at 10% level.

Although the results obtained in the pooled regressions are seemingly reasonable, it is likely that the coefficients are biased due to endogeneity. The endogeneity results from the potential omission of unobserved factors that are likely to be correlated with independent variables included in the model. Therefore, we need a model that some how corrects for this endogeneity. This is achieved, at least partially, in fixed
effects model by subtracting out the unobserved fixed effects by demeaning or first differencing the data.

Table 1.3 presents results from the estimation of the fixed effects model in logarithmic form. When the source country level unobserved time-invariant heterogeneity is accounted for, only the income levels in source countries and educational achievement of source country population retain the statistical significance. Not surprisingly, the coefficient on proportion of internet users also becomes statistically significant. The results in Table 1.3 indicate that for every 1% increase in per capita income at the source country, there is a decrease of 1.36% in the potential migration rate. Similarly, for every one percent increase in population with at least secondary education, the potential migration rate decreases by 1.16%. This is a slightly puzzling finding because most previous empirical studies of the determinants of international migration have found that more educated people are more likely to migrate. One potential explanation for this is that the aggregate level of educational attainment is just a proxy for overall quality of life in source countries. Therefore a rising educational attainment is indicating better quality of life at home and hence a lower drive to emigrate. It is also possible that the positive sign on educational achievement found in previously conducted studies, that do not distinguish between migration aspiration and actual migration, may be merely suggesting that educated individuals are more often able to actually convert their migration intentions into migration actions.

Finally, as access to internet improves by one percent, potential migration rate, as measured by US DV applications increases by 0.42%. According to the results generated in fixed effects model, income at destination, in this case the US, is not a driver of potential migration rate. This is possibly because most year-to-year variation in DV applications comes from developing countries whose per capita income is far below the income in US.
Table 1.2: Drivers of Potential Migration to the US: Results from Pooled Regression

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source per capita GDP</td>
<td>-1.283</td>
<td>0.434</td>
<td>0.004</td>
</tr>
<tr>
<td>US per capita GDP</td>
<td>2.013</td>
<td>2.568</td>
<td>0.435</td>
</tr>
<tr>
<td>Source political and civil liberties</td>
<td>0.099</td>
<td>0.110</td>
<td>0.371</td>
</tr>
<tr>
<td>Immigrant stock</td>
<td>0.676</td>
<td>0.255</td>
<td>0.010</td>
</tr>
<tr>
<td>Population between 15 and 29</td>
<td>-0.513</td>
<td>0.314</td>
<td>0.107</td>
</tr>
<tr>
<td>Population with at least secondary education</td>
<td>0.946</td>
<td>0.417</td>
<td>0.026</td>
</tr>
<tr>
<td>Internet users at source</td>
<td>0.115</td>
<td>0.253</td>
<td>0.649</td>
</tr>
</tbody>
</table>

Dep. var is potential migration rate (defined as the number of US DV applications per 1000 adult population), $N = 496$, $R^2 = 0.4017$. All variables except the political and civil liberties are in logarithmic form. Observations are weighted by average adult population in source countries and standard errors are clustered by country of origin.

Table 1.3: Drivers of Potential Migration to the US: Results from Fixed Effects Model

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source per capita GDP</td>
<td>-1.364</td>
<td>0.530</td>
<td>0.012</td>
</tr>
<tr>
<td>US per capita GDP</td>
<td>1.176</td>
<td>0.936</td>
<td>0.213</td>
</tr>
<tr>
<td>Source political and civil liberties</td>
<td>-0.016</td>
<td>0.098</td>
<td>0.866</td>
</tr>
<tr>
<td>Immigrant stock</td>
<td>0.318</td>
<td>0.208</td>
<td>0.130</td>
</tr>
<tr>
<td>Population between 15 and 29</td>
<td>1.108</td>
<td>1.140</td>
<td>0.334</td>
</tr>
<tr>
<td>Population with at least secondary education</td>
<td>-1.160</td>
<td>0.458</td>
<td>0.013</td>
</tr>
<tr>
<td>Internet users at source</td>
<td>0.420</td>
<td>0.192</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Dep. var is potential migration rate (defined as the number of US DV applications per 1000 adult population), $N = 496$, $R^2$ within = 0.2570. All variables except the political and civil liberties are in logarithmic form. Observations are weighted by average adult population in source countries and standard errors are clustered by country of origin.
1.6 Alternative Specifications

Some empirical studies of the determinants of international migration, such as Mayda (2010) and Mitchell, Pain, and Riley (2011), model international migration as a dynamic process positing that a lagged migration flow may have implications to current migration flow. A case for a dynamic specification of the willingness to migrate can be made on the grounds that willingness to migrate can have persistence arising from inertia in revising expectations about the destination countries or behavioral changes arising from past willingness to migrate (for example, working to make oneself able to migrate in future which reinforces the conditions that initially contributed in forming willingness to migrate). To see if such a dynamic specification significantly alters any of the results obtained in this paper, I specify a dynamic panel data model of the following type:

\[
pmr_{jt} = \alpha + \tau \times pmr_{j,t-1} + \beta_1 \times pcgdp \_source_{j,t} + \beta_2 \times pcgdp \_dest_{t} + \beta_3 \times average \_liberties_{j,t} \\
+ \beta_4 \times internet \_access_{j,t} + \beta_5 \times immigrant \_stock_{j,t} + \beta_6 \times age \_15to29_{j,t} \\
+ \beta_7 \times percent \_seceduc_{j,t} + \mu_j + u_{jt} \tag{1.7}
\]

where \( j = 1, 2, \ldots, N \); \( t = 2, 3, \ldots, T \) and \( \tau < 1 \).

The presence of a lagged dependent variable in the specification implies that simple demeaning process that corrects the problem of endogeneity in the case of fixed effects model cannot completely correct the problem here. That is because the mean of lagged dependent variable, \( pmr_{j,t-1} \), contains values of the dependent variable for time = 1 to \( T - 1 \); and the the mean of the error which is also subtracted from \( u_{jt} \) contains contemporaneous values of \( u_{jt} \) for \( t = 1 \) to \( T \) (Baum, 2013). The resulting
correlation between the lagged dependent variable and the error terms even after demeaning causes the so called Nickel bias. The problem is especially concerning when the time period $T$ is small and inclusion of additional regressors does not mitigate the problem.\textsuperscript{6}

Anderson and Hsiao (1982) proposed that one way to get around the above endogeneity issue is to use the instrumental variables that are present in the structure of the model itself. The process involves estimating the model in its first differences and using certain lags of dependent variable and other exogenous regressors as instrument for the endogenous term on the right hand side. Arellano and Bond (1991) pointed out that there is a larger number of instruments available to correct for endogeneity in GMM framework. Arellano and Bond (1991) suggest using not only the levels of certain lags of dependent variable but also the lagged differences of the dependent variable and lags of other pre-determined regressors. The resulting estimator is commonly referred to as system GMM.

Table 1.4 and 1.5 present respectively the GMM one step and two step estimates for the model specified in equation 1.7. The coefficients on source country income and destination country income are statistically significant in both sets of results. These results are somewhat different from those generated in fixed effects model but the lagged dependent variable is insignificant in both sets of results generated in dynamic specifications which indicates that the dynamic specifications are probably uncalled for in modelling potential migrate rate.

\textsuperscript{6}See Bond (2002) for an elegant explanation of this problem.
Table 1.4: Results from Dynamic Model: GMM One-step Estimates

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged Potential migration rate</td>
<td>0.369</td>
<td>0.230</td>
<td>0.109</td>
</tr>
<tr>
<td>Source per capita GDP</td>
<td>-2.836</td>
<td>1.465</td>
<td>0.053</td>
</tr>
<tr>
<td>US per capita GDP</td>
<td>4.737</td>
<td>42.301</td>
<td>0.040</td>
</tr>
<tr>
<td>Source political and civil liberties</td>
<td>-0.758</td>
<td>0.483</td>
<td>0.117</td>
</tr>
<tr>
<td>Immigrant stock</td>
<td>0.234</td>
<td>0.265</td>
<td>0.929</td>
</tr>
<tr>
<td>Population between 15 and 29</td>
<td>0.962</td>
<td>0.184</td>
<td>0.602</td>
</tr>
<tr>
<td>Population with at least secondary education</td>
<td>0.820</td>
<td>2.707</td>
<td>0.762</td>
</tr>
<tr>
<td>Internet users at source</td>
<td>0.497</td>
<td>0.388</td>
<td>0.201</td>
</tr>
</tbody>
</table>

Dependent variable is potential migration rate (defined as the number of US DV applications per 1000 adult population), N = 412, Number of instruments = 15. All variables except the political and civil liberties are in logarithmic form. Observations are weighted by average adult population in source countries and standard errors are clustered by country of origin.

Table 1.5: Results from Dynamic Model: GMM Two-step Estimates

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged potential migration rate</td>
<td>0.297</td>
<td>0.282</td>
<td>0.291</td>
</tr>
<tr>
<td>Source per capita GDP</td>
<td>-3.469</td>
<td>1.425</td>
<td>0.015</td>
</tr>
<tr>
<td>US per capita GDP</td>
<td>4.890</td>
<td>2.593</td>
<td>0.059</td>
</tr>
<tr>
<td>Source political and civil liberties</td>
<td>-0.716</td>
<td>0.099</td>
<td>0.000</td>
</tr>
<tr>
<td>Immigrant stock</td>
<td>0.722</td>
<td>0.833</td>
<td>0.386</td>
</tr>
<tr>
<td>Population between 15 and 29</td>
<td>-1.743</td>
<td>2.338</td>
<td>0.456</td>
</tr>
<tr>
<td>Population with at least secondary education</td>
<td>1.098</td>
<td>2.723</td>
<td>0.687</td>
</tr>
<tr>
<td>Internet users at source</td>
<td>0.614</td>
<td>0.512</td>
<td>0.231</td>
</tr>
</tbody>
</table>

Dependent variable is potential migration rate (defined as the number of US DV applications per 1000 adult population), N = 412, Number of instruments = 15. All variables except the political and civil liberties are in logarithmic form. Observations are weighted by average adult population in source countries and standard errors are clustered by country of origin.
1.7 Conclusion

The fact that every individual willing to migrate isn’t able to do so is widely recognized in the literature and is also supported by anecdotes and news reports. Although analysis of migration as a two step process has recently begun to address that distinction between potential and actual migrants, literature is in demand of conceptual clarity in modelling and estimation with good data.

This paper extends the emerging literature on determinants of potential migration by using administrative data from the US DV lottery. Not contrary to the previous literature, the analysis finds that higher income levels at the source decrease potential migration rate. However, contrary to suggestions in previous literature, a higher educational achievement of the source country population decreases potential migration rate.

The quantitative estimates of the driver of potential migrate rate obtained in the paper can be used to forecast migration flows in hypothetical scenarios of open borders. The results obtained in this chapter not only highlight the forces driving migration intentions but also add to our understanding for composition of the US labor force.
CHAPTER 2

SUPPLY OF IMMIGRANT ENTREPRENEURS AND NATIVE ENTREPRENEURSHIP

2.1 Introduction

More than 11% of all immigrant workers in the United States (US) operate their own businesses (Hipple, 2010). Immigrants not only lead natives in business creation rate and business ownership rate but also are more likely to hire employees (Fairlie, 2012). Moreover, immigrants constitute a group of highly innovative entrepreneurs. In the year 2006, for example, immigrants contributed to nearly one-fourth of all international patent applications filed from the US even though they made only 15.3% of the US labor force (Wadhwa, Saxenian, Rissing, & Gereffi, 2007).

Despite such a large proportion of immigrant workers engaging in entrepreneurial activities, the question of how immigrant entrepreneurs affect the entrepreneurial propensities of natives is not well understood. On one hand, immigrant entrepreneurs increase demand for intermediate goods and services, stimulate technological innovation (Duleep, Jaeger, & Regets, 2012), aid in cross-border technology diffusion (Kerr,
2008), and facilitate international trade (Gould, 1994). These effects are likely to help in growth of native-owned businesses (Coad & Rao, 2008) and encourage more natives to take up entrepreneurship (the “crowding in” effect). On the other hand, immigrant entrepreneurs compete with the native-run businesses by cutting costs and accepting lower profits (Ackerman & Tellis, 2001; Joona, 2011), potentially displacing natives from entrepreneurial activities (the “crowding out” effect). The opposing forces make the net impact of immigrant entrepreneurs on native entrepreneurial propensities theoretically ambiguous. This paper contributes to the literature by providing the first quasi-experimental evidence for the net impact of immigrant entrepreneurs on entrepreneurial propensities of natives.

The impact of immigrant entrepreneurs on native entrepreneurial propensities has a number of policy implications. Entrepreneurship is a significant component of economic growth (Haltiwanger, Jarmin, & Miranda, 2013; Rupasingha & Goetz, 2013; Van Stel, Carree, & Thurik, 2005). In the US alone, 28 million small businesses account for 54% of nationwide sales and provide 55% all jobs (Small Business Administration). Entrepreneurship is also a path out of poverty for many small business owners (Bruton, Ketchen, & Ireland, 2013). Due to these socio-economic benefits, governments at the federal, state, and local level provide tax breaks, subsidies, and other incentives to promote entrepreneurship (Rupasingha & Goetz, 2013). Moreover, many nonprofit foundations invest generously in entrepreneurial development (Goetz, Partridge, Deller, Fleming, et al., 2010). If immigrant entrepreneurs have a net positive effect on the entrepreneurial propensities of natives, a potential policy response would be to encourage further entrepreneurship among immigrants, for example by improving their access to credit or by issuing more entrepreneurship visas.¹ On the

¹Many immigrant receiving countries have implemented immigration policies that provide preferential treatment to potential business owners. While Canada, Australia and the United Kingdom explicitly provide special visas by the name of ‘start up visa’, ‘entrepreneur visa’, and ‘business talent visa’ respectively, the US provides ‘green card through investment’ and also has provision to allow permanent residence to potential entrepreneurs under categories like ‘national interest waiver’ and ‘exceptional ability in business’.

27
contrary, if immigrant entrepreneurs crowd out native entrepreneurs, the appropriate policy response hinges on the quality of the businesses immigrants create.\textsuperscript{2}

Previous literature on immigration and entrepreneurship is limited to either describing immigrant entrepreneurship or evaluating the impact of overall immigration on aggregate levels of entrepreneurship. This paper departs from the existing literature in that it estimates the direct causal impact of an exogenous change in the number of immigrant entrepreneurs on native entrepreneurial propensities. Specifically, I use data from the Annual Social and Economic (ASEC) Supplement of Current Population Survey (CPS) to study how within state variation in the supply of immigrant entrepreneurs affects the entrepreneurial propensities of natives. Econometrically, I model entrepreneurial propensity of natives as a function of the supply of immigrant entrepreneurs at the state level and other determinants of entrepreneurial tendency. An obvious concern with this approach is that the supply of immigrant entrepreneurs within a state is endogenous. Therefore, I use an instrumental variables strategy to circumvent the endogeneity concerns.

An important distinction between this paper and the previous literature lies in how entrepreneurship is defined. The standard approach in empirical studies of entrepreneurship is to consider all forms of self-employment as entrepreneurship (Blanchflower & Oswald, 1998). However, the emphasis on the difference between incorporated and unincorporated self-employment in the US context is growing (R. Levine & Rubinstein, 2013). Incorporating a business protects owners from the business liabilities at the cost of higher registration fees, extra paper work, and usually more complex legal compliance requirements (Suek, 2010). The higher cost associated with starting and maintaining an incorporated business works as a sorting mechanism to

\textsuperscript{2}For example, if immigrants create high quality businesses that make a net positive contribution to the economy even after accounting for potential losses due to crowding out, it may still be beneficial to promote entrepreneurship among the immigrants. However if immigrant-owned businesses are not necessarily better than those they displace, then steps to promote entrepreneurship among the immigrants or to attract immigrant entrepreneurs from abroad would be unwarranted.
attract only the individuals best suited to succeed as entrepreneurs. As a result, the incorporated self-employed individuals not only differ from the unincorporated self-employed in a wide range of personal characteristics such as ability, risk aversion, parental characteristics, teenage experience, etc. but also earn significantly different payoffs in the labor market (R. Levine & Rubinstein, 2013). To address the contrast between these entrepreneur types, I conduct separate analyses of the unincorporated self-employment and the incorporated self-employment.

There are many reasons to suspect heterogeneity in the impact of immigrant entrepreneurs. Often people with particular skills, educational attainment, age group, and ethnicity self-select into migration (Borjas, 1987b). Because the selection is driven by a number of socio-economic differences between source and destination countries, immigrant quality varies significantly across year-of-entry cohorts. However, on average, immigrant workers in the US are less educated than natives, have lower income, are slightly older, and are heavily concentrated in a few occupations such as farming, construction, cleaning and maintenance, food preparation, and food service (Camarota, 2012). The pre-migration selection based on skill and the post-migration occupational sorting of immigrants points to potentially heterogenous impact of immigrants on different groups of natives. For example, many low-skilled immigrants may complement high-skilled natives in the labor market but they are likely to compete with the low-skilled natives (Altonji & Card, 1991; Asali, 2013). Similarly, immigrants are concentrated in industries with less stringent licensing and language requirements due to educational and language mismatch (Boyd & Thomas, 2001), which makes the natives engaged in those industries more likely to be affected by immigrants. Therefore, I investigate the possibility of heterogenous impact of immigrant entrepreneurs across natives' skill levels and industries.

My findings suggest that, on average, unincorporated self-employed immigrants have no impact on the unincorporated self-employment propensities of natives. How-
ever, the incorporated self-employed immigrants crowd in natives into incorporated self-employment. Specifically, a 1% increase in the supply of incorporated self-employed immigrants increases the supply of incorporated self-employed natives by 0.11%. This increase is roughly equivalent to one additional native entrepreneur for every two immigrant entrepreneurs added to the US economy.

Moreover, I find that the crowding in observed in incorporated self-employment is mainly driven by the crowding in of high skilled natives. When the sample is segregated by industry, results indicate statistically significant crowding in of natives into incorporated self-employment in trade and transportation sectors. Furthermore, there is some indication that unincorporated self-employed immigrants engaged in manufacturing sector crowd out natives from unincorporated self-employment in that sector. These results highlight the multifaceted consequences of immigrant entrepreneurship and are suggestive of the appropriateness of skill-biased and industry-biased immigration policies.

The remainder of the paper is organized as follows. The next section provides a background on related literature. Section 3 discusses related economic theory and develops a conceptual framework necessary for econometric analysis. Section 4 discusses data. Section 5 lays out the estimation framework. Section 6 presents results. Finally, Section 7 presents robustness checks, and Section 8 concludes.

### 2.2 Related Literature

This paper is closely related to studies that investigate the impact of immigration on welfare of natives. Most previous studies, both theoretical and empirical, on the topic examine the impact of immigrant workers on wages and employment prospects of native workers. Conclusions from theoretical models in this class of studies depend on the assumptions placed on the relative skill structure of immigrant workers and
native workers (which ultimately determine the elasticity of substitution between the
two groups) and the elasticity of capital supply (Dustmann, Glitz, & Frattini, 2008;
Friedberg & Hunt, 1995; Manacorda, Manning, & Wadsworth, 2012). For example,
if capital supply is perfectly elastic and immigrant workers are perfect substitutes for
native workers, immigration does not affect the wages and employment prospects of
natives. However, if immigrant workers have different skill composition than native
workers, there will be winners and losers among natives even with perfectly elastic
capital supply.

Empirical studies on the topic are abundant (for example Card (1990), Borjas,
Freeman, Katz, DiNardo, and Abowd (1997), Winter-Ebmer and Zweimüller (1999),
Hu (2000), Ottaviano and Peri (2012), Borjas, Grogger, and Hanson (2011) to state
only a few) but report widely varying results. Nevertheless, a general consensus is
that in the long run immigrant workers, on average, neither lower wages of native
workers (Peri, 2014) nor displace them from their jobs (Constant, 2014). Increased
productivity, better skill mix, and high rate of entrepreneurship among immigrants
are some of the benefits of immigration highlighted in these studies.\(^3\)

Although immigration literature frequently discusses immigrant entrepreneurship
as an important benefit of immigration, credible economic research on how immigrant
entrepreneurs affect native entrepreneurship is lacking. However, a handful of studies
examine the closely related issue of how immigration affects entrepreneurship among
the natives. Light and Sanchez (1987), probably the earliest among them, utilizes
data collected from the 1980 public-use samples of U.S. Census to find that a higher
percentage of foreign-born in labor force results in a higher aggregate self-employment
rate. The study attributes up to 52% of increase in non-farm self-employment rate
in the 80s to the immigration influx of the 70s.

More recently, Fairlie and Meyer (1997) use census data from 1980 and 1990 to

\(^3\)See Kerr (2013) for a survey of recent empirical studies on the economic impacts of immigration.
estimate the impact of immigration on African American self-employment and find no significant effect. However, in a subsequent related study, the authors look at the impact of immigration on the self-employment propensities of the non-African American population in the US and find a small negative impact (Fairlie & Meyer, 2003).

Hunt and Gauthier-Loiselle (2010) examine the impact of skilled immigration on innovation, measured by patenting behaviour, in the US and find that a one percentage point increase in the share of immigrant college graduates in the population increases patents per capita by 6%. Kerr and Lincoln (2010) examine the impact of high-skilled immigrants, entering on H-1B visas, on US technology formation. The authors conclude that total innovation, measured by patenting and employment in science and engineering, increases with higher H-1B admission levels.

Although important, these studies investigate the impact of immigration on post-immigration aggregate levels of entrepreneurship or the entrepreneurship of the natives. They fall short of disentangling the impact of immigrant entrepreneurs from that of immigrants in general. Since a wave of immigration adds both producers and consumers of goods and services to the economy, the impact of immigration is a result of the economic activities of the immigrant entrepreneurs and the non-entrepreneurs. Isolating the impact of immigrant entrepreneurs from that of immigration not only informs policies related to immigration and entrepreneurship development but also furthers our understanding of the determinants of entrepreneurship among natives.

### 2.3 Conceptual Framework

Economic theory suggests several mechanisms through which immigrant entrepreneurs may impact native entrepreneurial outcomes. Immigrants often create novel businesses by innovating or facilitating cross-border technology diffusion (Kerr, 2008).
This increases the entrepreneurial propensities of the natives by working through the channel of knowledge spillover (Acs, Braunerhjelm, Audretsch, & Carlsson, 2009). Increased demand for goods and services that are inputs in immigrant-owned businesses is likely to encourage more natives to enter entrepreneurship. Moreover, immigrants’ knowledge of international markets helps reduce trade-related transaction costs leading to increased exports and imports (Gould, 1994), which further strengthens the crowding in effect.

However, many immigrants compete with natives for similar entrepreneurial opportunities. Such immigrants may hurt chances of native individuals becoming entrepreneurs by seizing the business opportunities that would otherwise go to natives (Fairlie & Meyer, 2003). The crowding out effect is potentially stronger if immigrants are willing to accept lower profits or are able to cut costs of production. Immigrant workers tend to have lower reservations wages due to different frames of references (Constant, Krause, Rinne, & Zimmermann, 2010); immigrant business owners may be willing to accept lower profits for similar reasons. Immigrant entrepreneurs are able to slash costs of production if they have access to cheaper inputs. For instance, when wages are determined by Nash bargaining between employer and employees, immigrant entrepreneurs are able to extract higher rents from immigrant workers if they are informed of immigrant workers’ lower reservation wages. Similarly, immigrant-owned businesses are able to pay lower wages if they can more easily hire undocumented immigrants, a group characterized by low labor supply elasticity.

While the preceding theoretical discussion suggests that immigrant entrepreneurs may have substantial impact on the entrepreneurial propensities of natives, it is difficult to sign this impact, ex ante, because of the opposing forces working via multiple channels. Therefore, to estimate the impact empirically, I build up on the conceptual framework of self-employment outlined by Borjas (1986). In this econometric setup, an individual $i$ chooses self-employment based on the difference between market wage
If we define an index function $I$ such that

$$I_i = y_i - w_i = C_i \pi + v_i$$  \hspace{1cm} (2.1)

where $C_i$ is the vector of socio-economic characteristics which affect $y_i$ and $w_i$, $v_i$ is the disturbance term, and $\pi$ is a parameter vector, then the individual chooses self-employment when $I_i > 0$. The probability of an individual being self-employed is then given by

$$P_i = Pr[I_i > 0|C_i] = Pr[v_i > -C_i \pi|C_i]$$ \hspace{1cm} (2.2)

If $v_i$ is distributed symmetrically, then

$$P_i = Pr[v_i < C_i \pi|C_i] = F(C_i \pi)$$ \hspace{1cm} (2.3)

where $F$ is the cumulative distribution function of the random variable $v_i$. The vector $\pi$ can now be estimated in a standard regression framework by specifying the functional form of $F$.

It is difficult to specify, with certainty, the exact set of variables that goes into the vector $C$. However, the theoretical discussion suggests that, in addition to other ordinarily considered micro and macro level determinants of labor market outcomes, the supply of immigrant entrepreneurs plays an important role in a native individual’s choice of self-employment. For our purpose, this implies a reduced form model which specifies the self-employment propensity of natives as a function of supply of immigrant entrepreneurs and other relevant controls.

### 2.4 Data and Variables

The main source of data for this study is the Annual Social and Economic Supplement (ASEC), popularly known as the March supplement, of the Current Population Survey.
(CPS). I use the uniform extracts of the CPS prepared by the Center for Economic Policy Research (CEPR). The CPS is administered jointly by the US Census Bureau and the US Bureau of Labor Statistics (BLS) and is the primary source of labor force statistics for the US population. While the basic CPS is a monthly survey of a nationally representative sample of approximately 60,000 households, the ASEC supplement provides data from an additional questionnaire administered every March. In addition to the usual demographic and labor force data, the ASEC supplement provides a rich set of information, on work experience, income, noncash benefits, and migration.

The data used in the current analysis span the period between 1996 and 2006. I focus on data after 1996 because the CPS underwent a significant redesign in the year 1996 (BLS). In the main analysis, I do not include data after 2006 for two reasons. First, the ‘necessity entrepreneurship’ rates increase for both immigrants and natives during recessions when the unemployment rate is high (Constant & Zimmermann, 2004; Figueroa-Armijos, Dabson, & Johnson, 2012). Including these years can lead to mistakenly identifying the recession-effects as positive effects of immigrant entrepreneurs on native entrepreneurial propensities. Second, there are concerns that recessions have potentially different effect on the labor market behavior of natives and immigrants (Constant & Zimmermann, 2004; Goldstein & Peters, 2014) which means the problem cannot be solved by simply including a dummy variable for recession years.

The analysis is limited to labor force participants in nonfarm households who are between the ages of 16 and 64. Several other sample restrictions are made for various sub-sample analyses. After employing listwise deletion to retain only records with complete information on variables necessary for this analysis, the pooled sample has annual observations on 732,096 individuals. Among them, 621,960 are natives and

Listwise deletion is problematic if missing values systematically differ from the completely observed cases. Because I control for a large number of demographic variables in the regressions,
110,136 are immigrants.

The CPS data are well suited for this study due to sizable immigrant population surveyed in them. While longitudinal data sets like the National Longitudinal Survey of Youth (NLSY) or the Panel Study of Income Dynamics (PSID) would normally be preferable for the analysis of dynamic changes in entrepreneurial activity of individuals over time, their small sample sizes and particularly smaller percentage of immigrants in the sample, make them less desirable for this analysis. Although the CPS does not track same individuals over the 11 year period considered in this study, the repeated cross-sectional nature of the data provides opportunity to capture aggregate trends in entrepreneurship.

It is important to remember that the CPS data have their own share of limitations, such as having a considerable amount of noise in identifying occupational and industry switches and providing a poor measure of annual occupational mobility (Kambourov & Manovskii, 2013). However, they have been extensively used in the analysis of the US labor market, including topics such as the impact of immigration on wages of natives and the impact of immigration on self-employment of natives (Fairlie & Meyer, 2003; Orrenius & Zavodny, 2007; H. J. Lee & Tomohara, 2008; Chassamboulli & Palivos, 2014).

### 2.4.1 Outcome variable

The outcome variable of interest in this analysis is the entrepreneurial propensity of natives. It is a binary variable indicating whether a native individual is an entrepreneur. Exactly what constitutes entrepreneurship is far from settled in the literature. Definitions range from an entrepreneur being an individual ‘who comes up

and the sample is remarkably large even after deletion, it is likely that missing values are missing at random. Note, however, that the data is subject to some imputation by the CPS before it is released. Exact details on the imputation are available at [http://www.census.gov/cps/methodology/unreported.html](http://www.census.gov/cps/methodology/unreported.html).
with new combinations of means of production’ to anyone who ‘sees a profit oppor-
tunity and creates an organization’ to pursue it (Gedeon, 2010). Although, past empirical studies have considered all forms of self-employment as entrepreneurship (Blanchflower & Oswald, 1998), more recent studies have highlighted the difference between the incorporated and the unincorporated self-employment in the US context (Hipple, 2010; R. Levine & Rubinstein, 2013; Olds, 2013). The sharp differences between two types of self-employment, discussed in Section 1, mandate separate econometric analysis of the incorporated self-employment and the unincorporated self-employment. To that end, I construct separate indicators for incorporated and unincorporated self-employment as observed in the March CPS.

2.4.2 Key explanatory variable

The key explanatory variable in this paper is the state-level supply of immigrant entrepreneurs, which is the number of immigrant entrepreneurs operating a business in a state in a particular year. For each state, this variable is constructed by counting the number immigrants reporting self-employment in a given year. For distinct analysis of incorporated and unincorporated self-employment, I construct separate measures for the supply of incorporated immigrant entrepreneurs and unincorporated immigrated entrepreneurs. In constructing these state level aggregate measures, I use the same definition of self-employment as used in creating indicators for outcome variable.

2.4.3 Additional control variables

In addition to the key predictor, the regression models control for a number of individual characteristics and state characteristics potentially correlated with the outcome variables. The individual level controls include age, education, gender, marital status, race, number of children, city residence status, indicator for public health
insurance coverage, and indicator for health insurance coverage through spouse’s policy. These variables are theoretically and empirically important determinants of self-employment (Bailey, 2013; Simoes, Crespo, & Moreira, 2015; Robinson & Sexton, 1994; Zissimopoulos & Karoly, 2007). The state level controls include the annual unemployment rate, average annual wage earnings, and population. These variables together proxy for the state level socioeconomic conditions that influence the choice of self-employment (Fairlie & Meyer, 2003).

2.5 Methods

2.5.1 The baseline model

For each type of self-employment - i.e. unincorporated and incorporated - the baseline econometric model takes the form:

$$ SelfEmployed_{ist} = \beta_0 + \beta_1 \times ImmiEnt_{st} + Z'_{st} \beta_2 + X'_{ist} \beta_3 + \eta_s + \lambda_t + \gamma.t \nu_s + \epsilon_{ist} $$ (2.4)

where $SelfEmployed_{ist}$ indicates whether a native individual $i$ in state $s$ in the year $t$ is self-employed, $ImmiEnt_{st}$ is the the supply of immigrant entrepreneurs in state $s$ and year $t$, $X_{ist}$ and $Z_{st}$ are respectively the vectors of individual characteristics and state characteristics discussed in the preceding section, and $\beta$s are parameter vectors. $\eta_s$, $\lambda_t$, and $\gamma.t \nu_s$ are state fixed effects, year fixed effects, and state specific linear trends respectively. $\epsilon_{ist}$ is error term.

Including state fixed effects in the model implies that the within state variation in supply of immigrant entrepreneurs is used to identify the impact of immigrant entrepreneurship.
entrepreneurs. The state fixed effects are expected to capture all time invariant unobserved state characteristics that impact the entrepreneurial propensities. Similarly, year fixed effects are expected to capture all unobserved periodic shocks that affect the nation as a whole. The state specific linear trends control for differences in self-employment trends across states.

Because the outcome is a limited dependent variable I estimate Equation 2.4 with a linear probability model (LPM). In regression frameworks that use survey data to explain a micro-level dependent variable with a group level independent variable, standard errors can be biased downwards if the correlation of errors within the groups is not accounted for (Moulton, 1990; Bertrand, Duflo, & Mullainathan, 2004). One way to deal with the intra-group correlation is to use cluster robust variance estimator (CRVE) in hypothesis testing. Therefore, to ensure that correlation within clusters is sufficiently accounted for, I adjust all standard errors for clustering at the state level.

Under the assumptions that regressors in Equation 2.4 are exogenous, parameter \( \beta_1 \) estimates the impact of the supply of specific type of immigrant entrepreneurs on the specific type of entrepreneurial propensity of natives. For instance, when \( \text{SelfEmployed}_{ist} \) is the indicator for whether a native individual is self-employed in an incorporated business, a positive (negative) \( \beta_1 \) implies that incorporated immigrant entrepreneurs crowd in (crowd out) natives into (from) incorporated businesses. Similarly when \( \text{SelfEmployed}_{ist} \) represents self-employment in unincorporated business, a positive (negative) \( \beta_1 \) implies that unincorporated immigrant entrepreneurs, on average, crowd in (crowd out) natives into (from) unincorporated business.

The estimates of \( \beta_1 \) obtained from Equation 2.4 are unbiased and consistent only

---

6It is common in literature to estimate choice of self-employment by logit or probit models. I employ LPM for two reasons. First the maximum likelihood estimation techniques occasionally face convergence problem. Second, the instrumental variable models in the case of nonlinear models are not well-understood. As argued by Wooldridge (2010) and Angrist and Pischke (2008), in binary response models, LPM does a good job of approximating the average marginal effects of the explanatory variables. Nevertheless, in robustness checks, I compare the LPM an probit estimates obtained from the baseline model and find that they yield similar results.
if the supply of immigrant entrepreneurs is uncorrelated with the error term conditional on state characteristics, individual characteristics, the state fixed effects, and the year fixed effects. However, these assumptions can be invalid for many reasons. For instance, if native individuals avoid or select states with large supply of immigrant entrepreneurs (endogenous sorting), the estimates from LPM can be biased and inconsistent. Additionally, Equation 2.4 does not control for inter-state variation in the ease of doing business because no such data is available. If this variable is not constant over time, a remarkably plausible scenario, the estimates of $\beta_1$ will be upward biased.\footnote{This signing of bias is true under the assumption that the omitted variable positively affects the supply of immigrant entrepreneurs and the entrepreneurial propensity of the natives.}

An additional source of omitted variable bias is the inability to control for whether natives reside in ethnic enclaves. While ethnic enclaves may help immigrants succeed in labor market (Damm, 2009), natives may be less likely to operate businesses in immigrant enclaves. Not controlling for this potentially relevant factor causes downward bias in the estimate of $\beta_1$.

Another source of bias in the estimates of $\beta_1$ obtained from Equation 2.4 is the possible measurement error in the supply of immigrant entrepreneurs. The CPS over samples certain households such as those with children, minorities, and people of Hispanic origin (Passel, Suro, & Center, 2005). If the self-employment rate among Hispanics differs significantly from that of other immigrant groups, then over representation of Hispanics in the ASEC introduces error in the measurement of the supply of immigrant entrepreneurs. Measurement error in the outcome variable and the key variable of interest can also result from reporting errors, random or otherwise, common in survey data. Except in the simplest case of classical measurement error (CME), the sign of bias resulting from this source cannot be signed (Hyslop & Imbens, 2001).

In addition to these potential source of biases, endogeneity can result from reverse causality from the outcome variable to the regressor of interest. For example,
while it is true that immigrant entrepreneurs affect native entrepreneurial propensi-
ties, it is also likely that entrepreneurial propensity of natives in a state influences
the supply of immigrant entrepreneurs in that state.

2.5.2 The instrumental variable model

To address the above empirical concerns, I estimate two stage least squares (TSLS)
models. In TSLS models, a suitable instrumental variable enables causal identification
by acting as an exogenous source of variation in the endogenous regressor.

Assuming monotonicity in its effect, an instrument must satisfy two conditions to
ensure consistency of the estimates of $\beta_1$. First, the instrument must have sufficient
correlation with the supply of immigrant entrepreneurs after partialling out the effect
of other covariates (the relevance requirement) and second, it should not be correlated
with the error term $\epsilon$ in Equation 2.4 after conditioning on other covariates (the
exclusion restriction). Intuitively, we require an exogenous source of variation in
the supply of immigrant entrepreneurs which does not affect native self-employment
except through the channel of supply of immigrant entrepreneurs.

The quasi experiment provided by the implementation of the State Children’s
Health Insurance Program (SCHIP) across states and years provides such variation.
To understand how SCHIP induces a plausibly exogenous variation in the supply of
immigrant entrepreneurs, one needs to go back to the 1996 welfare reform. As the
welfare program Aid to Families with Dependent Children (AFDC) drew increasing
criticism in the 80s and 90s for ‘disincentivizing work’ and ‘promoting welfare depen-
dence’ among the beneficiaries, the US passed the Personal Responsibility and Work
Opportunity Reconciliation Act (PRWORA) in 1996. Two most important goals of
the Act were to encourage work among welfare recipients and achieve a significant
decline in welfare caseloads (Clinton, 1996). To that end, PRWORA replaced AFDC
program with the Temporary Assistance for Needy Families (TANF) and imposed
stricter work requirements for welfare eligibility. Additionally, the Act banned im-
migrants who have been in the US for less than five years, henceforth newly-arried
immigrants, from receiving federally funded means tested benefits such as TANF and
Medicaid.

In 1997, the Congress passed SCHIP to cover uninsured children living in low-
income families that make too much to qualify for Medicaid but too little to afford
private coverage. The program was rapidly implemented across states. By 2000,
every state including the District of Columbia (DC) had some form of the SCHIP.
The SCHIP promised a partnership between the federal and the state governments
and gave states flexibility in designing their own eligibility requirements. Taking
advantage of this autonomy, 15 states including the DC, henceforth ‘generous states’,
chose to include newly-arrived immigrants in their SCHIP. However, newly arrived
immigrants in other ‘less generous states’ remained ineligible for all means tested
public health insurance. Due to the provision of the extra safety net, referred to as
‘generous SCHIP’ hereafter, the supply of self-employed immigrants increased by up
to 28% in those 15 states (Olds, 2013). The generous SCHIP increased the supply
of immigrant entrepreneurs by acting through several mechanisms such as loosening
liquidly constraint and reducing the risk associated with starting a business (more on
this later). Because it seems unlikely that SCHIP coverage for immigrant children
would directly impact the entrepreneurial propensities of the natives, the availability
of generous SCHIP for immigrants makes a potentially good instrument for the supply
of immigrant entrepreneurs.

Specifically, the instrument takes the form of a binary variable that indicates the
availability of generous SCHIP in a given state and a given year (\(GenSCHIP_{st}\)).

---

8These states also provided Medicaid coverage to a select group of newly arrived adult immigrants
who met eligibility criteria based on disability, age, etc. (Fortuny & Chaudry, 2011). In 2009, an
amendment to the Children’s Health Insurance Program Reauthorization Act (CHIPRA) allowed
all states to provide medical coverage to lawfully present pregnant women and children without a
waiting period.
The information required to construct the instrument, i.e. the information on dates of SCHIP implementation and the information on states choosing to cover the children of newly arrived immigrants, comes from Rosenbach et al. (2001) and Olds (2013). Because the states implemented SCHIP at different times between 1997 and 2000, the policy provides within state variation generated in different years.

There are many reasons for which the provision of SCHIP to newly arrived immigrants can increase the supply of immigrant entrepreneurs. First, in the US, adults are often compelled to work for large employers to ensure continued access to affordable insurance for themselves and their kids (Fairlie, Kapur, & Gates, 2011). Some of these adults may switch into self-employment when there is a guaranteed access to publicly funded children’s health insurance. Second, public health insurance helps families save on premiums that would otherwise be paid for private coverage. Savings accumulated over time because of public coverage may ease liquidity constraint, a commonly cited obstacle to business creation (Evans & Jovanovic, 1989; Fairlie & Krashinsky, 2012). Third, the supply of immigrant entrepreneurs may increase when adults in families without any insurance coverage, who had previously chosen to remain out of labor force for child care, decide to enter the labor force due to reduced risk of leaving a child at day care or with a family member.

The validity of generous SCHIP (GenSCHIP) as an instrument depends on its excludability from Equation 2.4. In other words, generous SCHIP should have no direct impact on the self-employment propensities of natives. While it is not possible to directly test this assumption in just identified models, in results section I provide ample suggestive evidence which indicates that the assumption is reasonable.
2.6 Results

2.6.1 Descriptive statistics

I begin with close examination of the summary statistics. In the pooled sample that includes both the natives and the immigrants sampled during 1996-2006 period, the overall self-employment rate, defined as the proportion of self-employed individuals in non-agricultural labor force, is approximately 9.0%. Of the 9.0%, approximately 5.8% are self-employed in unincorporated businesses and only 3.2% are self-employed in incorporated businesses. That a significantly smaller number of entrepreneurs chose to incorporate their businesses, in spite of a number of legal benefits including the advantage of limited liability, indicates that incorporated self-employed individuals may indeed be different from unincorporated self-employed in certain aspects.

Table 2.1 and Table 2.2 present the summary characteristics of the self-employed natives and immigrants respectively. The statistics, calculated separately for the incorporated and the unincorporated, demonstrate that incorporated entrepreneurs differ from the unincorporated in many ways. The incorporated self-employed individuals are, on average, older - therefore, with potentially longer labor market experience - and more educated than the unincorporated self-employed. Similarly, incorporated self-employed are more likely to be married, have more children, and are more likely to live in metropolitan areas than their unincorporated counterparts. Women and minorities (Blacks and Hispanics) have a better representation in unincorporated than incorporated businesses. These differences further support the distinct treatment of the two types of self-employment.
Table 2.1: Summary Statistics: Self-employed Natives

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. deviation</td>
</tr>
<tr>
<td>Age</td>
<td>44.11</td>
<td>10.60</td>
</tr>
<tr>
<td>Female</td>
<td>0.39</td>
<td>0.48</td>
</tr>
<tr>
<td>Married</td>
<td>0.69</td>
<td>0.46</td>
</tr>
<tr>
<td>Children in household</td>
<td>0.82</td>
<td>1.15</td>
</tr>
<tr>
<td>City residence status</td>
<td>0.79</td>
<td>0.40</td>
</tr>
<tr>
<td>Less than high school</td>
<td>0.07</td>
<td>0.26</td>
</tr>
<tr>
<td>High school</td>
<td>0.31</td>
<td>0.46</td>
</tr>
<tr>
<td>Some college</td>
<td>0.29</td>
<td>0.45</td>
</tr>
<tr>
<td>College degree</td>
<td>0.19</td>
<td>0.39</td>
</tr>
<tr>
<td>Advanced degree</td>
<td>0.11</td>
<td>0.32</td>
</tr>
<tr>
<td>White</td>
<td>0.87</td>
<td>0.33</td>
</tr>
<tr>
<td>Black</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.04</td>
<td>0.20</td>
</tr>
<tr>
<td>Public health insurance</td>
<td>0.06</td>
<td>0.23</td>
</tr>
<tr>
<td>Covered through spouse’s insurance</td>
<td>0.34</td>
<td>0.47</td>
</tr>
<tr>
<td>Observations</td>
<td>35,897</td>
<td>20,325</td>
</tr>
</tbody>
</table>

*Note: Sample weights applied.*
Table 2.2: Summary statistics: Self-employed Immigrants

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. deviation</td>
</tr>
<tr>
<td>Age</td>
<td>43.05</td>
<td>10.26</td>
</tr>
<tr>
<td>Female</td>
<td>0.35</td>
<td>0.47</td>
</tr>
<tr>
<td>Married</td>
<td>0.73</td>
<td>0.44</td>
</tr>
<tr>
<td>Children in household</td>
<td>1.06</td>
<td>1.22</td>
</tr>
<tr>
<td>City residence status</td>
<td>0.96</td>
<td>0.19</td>
</tr>
<tr>
<td>Less than high school</td>
<td>0.07</td>
<td>0.26</td>
</tr>
<tr>
<td>High school</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>Some college</td>
<td>0.18</td>
<td>0.38</td>
</tr>
<tr>
<td>College degree</td>
<td>0.18</td>
<td>0.38</td>
</tr>
<tr>
<td>Advanced degree</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>White</td>
<td>0.28</td>
<td>0.45</td>
</tr>
<tr>
<td>Black</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>Public health insurance</td>
<td>0.05</td>
<td>0.22</td>
</tr>
<tr>
<td>Covered through spouse’s insurance</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>Observations</td>
<td>6,303</td>
<td></td>
</tr>
</tbody>
</table>

*Note*: Sample weights applied.
Next, before analyzing the individual level data, I examine how the outcome variable and the key explanatory variable vary at the state level. This gives us a picture of the relationship between native self-employment rate and immigrant entrepreneurs before we control for potential confounding variables. Figure 2.1 plots the native self-employment rate for each state-year cell against the corresponding supply of immigrant entrepreneurs. The raw data shows a negative relationship between the unincorporated entrepreneurial propensity of natives and supply of unincorporated immigrant entrepreneurs. The relationship is very weakly positive in case of incorporated entrepreneurship.
2.6.2 Baseline results

Figure ?? provides some suggestive evidence for the relationship between the native entrepreneurial propensities and the supply of immigrant entrepreneurs, however, it overshadows the effect of numerous demographic and economic changes driving the two variables. To estimate the causal impact of immigrant entrepreneurs on native entrepreneurship, I next estimate regression models that control for such factors.

Selected results generated in the baseline model are presented in Table 2.3. Only coefficients on immigrant entrepreneurs (ImmiEnt) are shown for brevity. All models control for individual characteristics, state characteristics, state fixed effects, year fixed effects, and the state specific linear trends. All coefficients in Table 2.3 are positive but only the estimate obtained in case of incorporated entrepreneurship is statistically significant at conventional levels. These estimates suggest that while unincorporated immigrant entrepreneurs have no significant impact on native self-employment, the incorporated immigrant entrepreneurs crowd in natives into incorporated self-employment. Specifically, a 1% increase in the number of incorporated immigrant entrepreneurs increases the probability of a native individual’s self-employment in an incorporated business by 0.0028 percentage points. Given that only 3.23% of natives in the sample are incorporated entrepreneurs, this increase is equivalent to an increase of approximately 0.09% in the number native entrepreneurs operating incorporated businesses.
Table 2.3: Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives: Linear Probability Model

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entrepreneurial propensity of natives:</td>
<td>0.0576</td>
<td>0.0323</td>
</tr>
<tr>
<td>Immigrant entrepreneurs</td>
<td>-0.0004</td>
<td>0.0028***</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>Observations</td>
<td>621,960</td>
<td>621,960</td>
</tr>
</tbody>
</table>

Note: Dependent variable is binary indicator for being an entrepreneur. The supply of immigrant entrepreneurs is log transformed. All models control for demographic characteristics (age, marital status, education, race, number of children, city residence status, public health insurance coverage and health insurance coverage through spouse’s insurance), time varying state characteristics (annual unemployment rate, average annual wage earnings, population), state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors adjusted for clustering at the state level are reported in parentheses. *p<.10. **p<.05. ***p<.01
As discussed in Section 4, the estimates from LPM obtained with potentially endogenous regressor, immigrant entrepreneurs (ImmiEnt), on the right hand side are likely to be biased and inconsistent due to simultaneity, reverse causality, measurement error, and omitted variables. The following subsection presents the results obtained from instrumental variable models that are expected to yield unbiased and consistent estimates.

2.6.3 Results from instrumental variable model

A challenge in establishing the causal relationship between the supply of immigrant entrepreneurs and the entrepreneurial propensities of natives is the possible endogeneity of the supply of immigrant entrepreneurs. In presence of an endogenous regressor, instrumental variables models can provide consistent estimates of the causal impact. However the identification strategy requires locating a suitable instrument. Specifically, the IV must be strongly correlated with endogenous regressor and uncorrelated with the error term. Thus, before presenting the IV estimates, I discuss results from the tests designed to test the suitability of the instrument.

Instrument strength

First, I conduct series of tests to check if the instrument is sufficiently correlated with the endogenous regressor. To this end I estimate first stage regressions for each entrepreneurship type. Estimates from the first stage regressions of the TSLS models presented in Table 2.4 show that the instrument (generous SCHIP) is strongly correlated with the supply of both unincorporated and incorporated immigrant entrepreneurs. After controlling for other independent variables, the provision of SCHIP coverage for the children of newly arrived immigrants is associated with 26.99% increase in the supply of unincorporated immigrant entrepreneurs and 21.22% increase
Table 2.4: First-stage Regressions: Association between Generous SCHIP and Supply of Immigrant Entrepreneurs

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average supply of immigrant entrepreneurs</td>
<td>23,383</td>
<td>13,030</td>
</tr>
<tr>
<td>Generous SCHIP</td>
<td>0.2699*** (0.0838)</td>
<td>0.2122*** (0.0540)</td>
</tr>
<tr>
<td>F-statistic</td>
<td>10.37</td>
<td>15.41</td>
</tr>
<tr>
<td>Prob &gt; F</td>
<td>0.0023</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Observations 621,960

Note: Dependent variable is natural logarithm of the supply of immigrant entrepreneurs. All models control for demographic characteristics (age, marital status, education, race, number of children, city residence status, public health insurance coverage and health insurance coverage through spouse’s insurance), time varying state characteristics (annual unemployment rate, average annual wage earnings, population), state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors adjusted for clustering at the state level are reported in parentheses. *p<.10. **p<.05. ***p<.01

in the supply of incorporated immigrant entrepreneurs. Given that approximately 6% of immigrants are unincorporated entrepreneurs and about 3 % are incorporated entrepreneurs, the first stage results indicate increases of approximately 1.6 and 0.6 percentage points in entrepreneurship rates of each type. The first stage coefficients on the instrument are strongly statistically significant with p values less than 0.01 in both regressions. These estimates are comparable to those found by Olds (2013) and they indicate that the instrument is strongly correlated with the supply of each type of immigrant entrepreneurs. Moreover, the F-statistics for joint significance of instruments in the first stage regressions for unincorporated and incorporated entrepreneurship are 10.37 and 15.41 respectively. Thus the instrument passes the ‘rule of thumb’ test recommended by Stock and Yogo (2005).

9Fairlie et al. (2011) estimate the impact of Medicare on business ownership rates among people older than 65 and find similar results.
Instrument validity

Because the error itself is unobserved, there are no straightforward econometric tests to establish that the chosen instrument is uncorrelated with the error term. Nevertheless, we can check, with reasonable reliability, if the instrument affects the outcome variable via channels other than the endogenous regressor (Angrist & Pischke, 2008).

There are at least two pathways through which immigrant children’s access to state-funded health insurance may directly affect self-employment decisions of the native individuals, thereby violating the exclusion requirement on the instrument. First, if immigrant children’s access to publicly funded health insurance crowds out native children from public insurance, which in turn influences the native self-employment, then TSLS estimates can be inconsistent.

To test if there was a crowding out of native children from public health insurance, I compare the insurance coverage rates of native children in the generous and the less generous states before and after the SCHIP implementation by using a difference-in-differences (DD) estimation strategy. Following the SCHIP implementation, if native children in the generous states experienced a significantly lower increase in coverage rates compared to those in less generous states, that would be an evidence in favor of the crowding out hypothesis.

I estimate the impact of generous SCHIP on two different measures of insurance coverage among native children. In the first model the dependent variable is an indicator for whether a native child has any insurance coverage, and in the second it is whether the child has Medicaid or SCHIP coverage. The results presented in Table A.1 do not indicate any crowding out of native children from Medicaid or SCHIP coverage in the generous states.

Second, the native individuals might engage in endogenous sorting behaviour in which they move to or away from generous states. This happens if states’ generosity

---

10See Section A.1 in Appendix for specific difference-in-difference (DD) model.
toward immigrants translates into more or less generosity toward natives. For example, generous states covering newly arrived immigrants may have generous income eligibility requirements making these states more attractive for natives at the margin. On the contrary, some states providing coverage for newly arrived immigrants may be offsetting the funds by providing coverage to fewer natives by having more stringent income eligibility guidelines.

To examine if the generous states differ significantly from less generous states in terms of generosity toward natives, I look at the SCHIP eligibility guidelines across states. States provide SCHIP coverage for children if the family income is below a certain percent of Federal Poverty Line (FPL). As of July 2006, the guidelines ranged from 140% of FPL to 350% of FPL (Ross, Cox, & Marks, 2007). To see the correlation between coverage for immigrants and overall generosity of SCHIP, I first categorize each state into a quintile in terms of eligibility guidelines, as shown in Figure ??, and generate an indicator variable to denote if a state falls in a particular quintile ranging from 1 to 5. The correlation (Pearson’s r) between this generosity indicator and provision of coverage for newly arrived immigrants is weakly positive at 0.28. This indicates that while, state generosity is somewhat related with immigrant coverage, this certainly is not a major concern.

While the generous SCHIP was probably not assigned randomly across the states, the results obtained in this paper remain valid as long as any non-additive unobservables in the error term $\epsilon$ of Equation 2.4 are not driving the policy. Furthermore, the economic literature has a long tradition of using state laws as exogenous treatments (Gruber & Madrian, 1997), (Leigh & Ryan, 2008), and (Boes, Marti, & Maclean, 2014).
In Table 2.5, I report the IV estimates of the impact of supply of immigrant entrepreneurs on the entrepreneurial propensities of natives. The TSLS estimates indicate that the impact of unincorporated immigrant entrepreneurs on the native entrepreneurial propensity of natives is statistically indistinguishable from zero. However, the incorporated immigrant entrepreneurs have a statistically significant positive impact on the incorporated self-employment propensity of natives. Although the estimates generated in TSLS models are qualitatively similar to the baseline estimates, they are slightly different in magnitude. The IV estimates imply that a 1% increase in the supply of incorporated immigrant entrepreneurs increases the probability of a native individual becoming an incorporated entrepreneur by 0.0036 percentage. This increase is equivalent to an increase of 0.11% in the number of native entrepreneurs operating incorporated businesses. The IV estimates are consistent with the potential omitted variable bias discussed in Section 4. It is evident that there is a net downward bias in the LPM estimate of $\beta_1$ in the case of incorporated entrepreneurship. The negative but statistically insignificant coefficient obtained in case of unincorporated entrepreneurship is suggestive of an upward bias in LPM estimates for that entrepreneurship type.

2.6.4 Heterogeneous impact of immigrant entrepreneurship

There are several reasons for which immigrant entrepreneurs may have heterogeneous impact across different groups of natives. Immigrants are a unique group of individuals often with limited language ability. They frequently suffer from skill mismatch in the host economy. Newly arrived immigrants also naturally lack extended networks of friends and family in the host country. Such limitations often force immigrants to be disproportionately concentrated in low-skilled jobs, at least prior to their cultural
Table 2.5: Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives: Two Stage Least Squares Model

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entrepreneurial propensity of natives:</td>
<td>0.0576</td>
<td>0.0323</td>
</tr>
<tr>
<td>Immigrant entrepreneurs</td>
<td>-0.0008 (0.0012)</td>
<td>0.0036** (0.0013)</td>
</tr>
<tr>
<td>Observations</td>
<td>621,960</td>
<td>621,960</td>
</tr>
</tbody>
</table>

Note: Dependent variable is binary indicator for being an entrepreneur. The supply of immigrant entrepreneurs is log transformed and instrumented by provision of generous SCHIP. All models control for demographic characteristics (age, marital status, education, race, number of children, city residence status, public health insurance coverage and health insurance coverage through spouse’s insurance), time varying state characteristics (annual unemployment rate, average annual wage earnings, population), state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors adjusted for clustering at the state level are reported in parentheses. *p < .10. **p < .05. ***p < .01

and economic assimilation in the host country. These limitations also raise the cost of acquiring information needed for business creation which forces immigrants to be concentrated in certain sectors that are characterized by low entry barriers such as low-startup capital requirement, minimal licensing requirements, easy skills transferability, and less stringent language requirements. For example, immigrants knowledge of ethnic food makes them more likely to be employed or self-employed in hospitality sector but strict licensing requirements can make foreign-educated lawyers, engineers, and doctors to work outside their area of expertise (Boyd & Thomas, 2001). Therefore, it is likely that groups of natives - such as high-skilled versus low-skilled or those engaged in construction versus retail trade for example - experience different impact from immigrant entrepreneurs.

**Heterogeneity across skill levels**

Table 2.6 presents the TSLS estimates of the impact of immigrant entrepreneurs on native entrepreneurial propensities when natives are separated by skill level. Only
Table 2.6: Heterogeneity in the Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives across Skill Levels: TSLS Estimates

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High Skilled</td>
<td>Low Skilled</td>
</tr>
<tr>
<td>Entrepreneurial propensity of natives:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0628</td>
<td>0.0556</td>
</tr>
<tr>
<td>Immigrant entrepreneurs</td>
<td>0.0033 (0.0027)</td>
<td>-0.0020 (0.0018)</td>
</tr>
<tr>
<td>First Stage F-stat</td>
<td>8.44***†</td>
<td>11.15***</td>
</tr>
<tr>
<td>Prob&gt;F</td>
<td>0.0054</td>
<td>0.0016</td>
</tr>
<tr>
<td>Observations</td>
<td>177,831</td>
<td>444,129</td>
</tr>
</tbody>
</table>

Note: Dependent variable is binary indicator for whether a native individual is an entrepreneur. The supply of immigrant entrepreneurs is log transformed. All models include control for time varying state characteristics (annual unemployment rate and average annual wage earnings, population), demographic characteristics (age, marital status, education, race, number of children, and city residence status), state fixed effects, year fixed effects, and state specific linear trends. Cluster robust standard error reported in parentheses. *p<.10. **<.05. *p<.01 †Instrument fails rule of thumb test for strength. However, estimates from limited information maximum likelihood are similar.

estimates from the preferred TSLS models are reported. High skilled natives are those who have a college degree and low skilled are those without a college degree. All estimates obtained in the case of unincorporated entrepreneurship are statistically indistinguishable from zero. However, results for incorporated entrepreneurship indicate that the crowding in observed in the case of incorporated self-employment is driven by the crowding in of the high-skilled natives. Specifically, a 1% rise in the number of incorporated immigrant entrepreneurs increases the probability of a high-skilled native individual becoming an incorporated entrepreneur by approximately 0.007 percentage points. Given that about 5.11% of high skilled natives operate incorporated businesses, this increase is equivalent to an increase of approximately 0.14% in the number of high skilled natives operating incorporated businesses.
Heterogeneity across industries

Immigrant entrepreneurs are mostly concentrated in retail trade, construction, transportation, finance, manufacturing and wholesale trade (See Figure ??).\textsuperscript{11} In Table 2.7, I report the TSLS estimates of the impact of immigrant entrepreneurs on entrepreneurial propensities of natives across these six industries.

All but one estimate for unincorporated entrepreneurship shown in the upper panel of Table 2.7 are statistically insignificant. The one statistically significant coefficient in the upper panel suggests that immigrant entrepreneurs operating businesses in manufacturing sector crowd out natives from unincorporated self-employment. Specifically, the coefficient implies that a 1% increase in the number of unincorporated immigrant entrepreneurs in manufacturing sector decrease the unincorporated self-employment propensity of natives in that sector by 0.0068 percentage points. This reduction is equivalent to a decrease of 0.17% in the number of natives operating unincorporated businesses in manufacturing sector.

The lower panel of Table 2.7 shows that the crowding in due to incorporated immigrant entrepreneurs is strongest in the retail and transportation sectors. Specifically, a 1% increase in immigrant entrepreneurs in retail trade increases the probability of natives’ incorporated self-employment in that sector by 0.078 percentage point. This increase in probability is equivalent to an increase of 0.52% in the number of incorporated native entrepreneurs in retail trade. Similarly, a 1% increase in number of immigrant entrepreneurs in transportation sector increases the probability of natives’ incorporated self-employment in that sector by 0.009 percentage points. This is equivalent to an increase of 0.25% in the number of native entrepreneurs operating

\textsuperscript{11}There are compatibility issues when using industry information over time in the CPS data. Details are available at http://www.bls.gov/cps cpsoccid.htm. The CEPR extracts provide consistent two digit codings so that meaningful comparisons can be made over time. In the CEPR extracts, the two digit codes classify jobs into 47 major groups for data before 2002 and 21 groups for data after that. Table A.2 presents classifications available for years before and after 2003. Figure ?? shows the 11 sectors that match from the two columns. I count all individuals engaged in non-matching sectors as ‘others’.

57
Figure 2.2: Distribution of Immigrant Entrepreneurs across Industries

Table 2.7: Heterogeneity in the Impact of Immigrant Entrepreneurs on Entrepreneurial Propensities of Natives across Industries: Two Stage Least Squares Model

<table>
<thead>
<tr>
<th>Industries</th>
<th>Retail</th>
<th>Construction</th>
<th>Transportation</th>
<th>Finance</th>
<th>Manufacturing</th>
<th>Wholesale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unincorporated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entrepreneurial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>propensity of natives:</td>
<td>0.1113</td>
<td>0.1920</td>
<td>0.0411</td>
<td>0.0775</td>
<td>0.0396</td>
<td>0.0226</td>
</tr>
<tr>
<td>Immigrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>entrepreneurs</td>
<td>0.0018</td>
<td>-0.0011</td>
<td>-0.0054</td>
<td>-0.0078</td>
<td>-0.0068**</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td>(0.0041)</td>
<td>(0.0115)</td>
<td>(0.0076)</td>
<td>(0.0302)</td>
<td>(0.0030)</td>
<td>(0.0106)</td>
</tr>
<tr>
<td>First stage F-statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p-value)</td>
<td>11.39</td>
<td>10.25</td>
<td>10.06</td>
<td>9.44†</td>
<td>7.44†</td>
<td>13.80</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0024)</td>
<td>(0.0026)</td>
<td>(0.0034)</td>
<td>(0.0088)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Observations</td>
<td>89290</td>
<td>43438</td>
<td>27401</td>
<td>47,252</td>
<td>78937</td>
<td>21,553</td>
</tr>
<tr>
<td>Incorporated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entrepreneurial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>propensity of natives:</td>
<td>0.1488</td>
<td>0.1636</td>
<td>0.0393</td>
<td>0.0951</td>
<td>0.0748</td>
<td>0.0601</td>
</tr>
<tr>
<td>Immigrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>entrepreneurs</td>
<td>0.0784*</td>
<td>0.0086</td>
<td>0.0099**</td>
<td>-0.0046</td>
<td>0.0010</td>
<td>-0.0020</td>
</tr>
<tr>
<td></td>
<td>(0.0416)</td>
<td>(0.0099)</td>
<td>(.0042)</td>
<td>(0.02549)</td>
<td>(0.0023)</td>
<td>(0.0049)</td>
</tr>
<tr>
<td>First stage F-statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p-value)</td>
<td>14.50</td>
<td>16.10</td>
<td>12.57</td>
<td>15.44</td>
<td>10.15</td>
<td>19.93</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0002)</td>
<td>(0.2601)</td>
<td>(0.0003)</td>
<td>(0.0025)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Observations</td>
<td>89290</td>
<td>43438</td>
<td>27401</td>
<td>47,252</td>
<td>78937</td>
<td>21,553</td>
</tr>
</tbody>
</table>

Note: Dependent variable is a binary indicator for whether a native individual is an entrepreneur. The supply of immigrant entrepreneurs is log transformed. All models include control for time varying state characteristics (annual unemployment rate and average annual wage earnings, population), demographic characteristics (age, marital status, education, race, number of children, and city residence status), state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors adjusted for clustering at the state level are reported in parentheses. †First stage F-statistic indicates weak instruments in these regressions. Estimates from limited information maximum likelihood (LIML) are similar. *p < .10. **p < .05. ***p < .01
incorporated businesses in transportation. Coefficients obtained for other sectors are statistically indistinguishable from zero.

Taken together, the estimates reported in Table 2.7 show that the incorporated immigrant entrepreneurs crowd in natives into incorporated self-employment in the retail and transportation sector. Moreover, unincorporated immigrant entrepreneurs in manufacturing may crowd out natives from unincorporated self-employment in that sector.

### 2.6.5 Economic significance and policy implications

The point estimates obtained in this analysis indicate economically significant impact of immigrant entrepreneurs. To put results in perspective, I calculate how adding one additional immigrant entrepreneur to the US economy affects the number of native entrepreneurs. In 2014, for example, there were 3,658,125 natives and 801,505 immigrants who were operating incorporated businesses in non-farm sectors.¹² The results generated in the TSLS models imply that adding two incorporated immigrant entrepreneurs, for instance by issuing two extra entrepreneurship visas, would result in one more native entrepreneur operating incorporated business. These calculations are suggestive of rather large economic gains from incorporated immigrant entrepreneurs.

Sectoral analysis shows that the crowding in effect of incorporated immigrant entrepreneurs is strongest in select sectors (i.e. trade and transportation among those analysed). However, there is also some evidence that immigrants operating unincorporated businesses in manufacturing may crowd out natives from unincorporated entrepreneurship in that sector. Given a significant role of entrepreneurship in economic growth and alleviation of poverty, these results point to a substantial benefit of a selective immigration policy that prioritizes visas for high skilled individuals in certain sectors. More precisely, an immigration policy that intends to take advantage

---

¹²Estimates based on 16-64 year old individuals surveyed in ASEC 2015.
of immigrant entrepreneurship should issue visas to potential entrepreneurs who are likely to operate high quality businesses in industries that promise the most spillover benefits. In the US context, one way to accomplish this is to issue entrepreneurship visas to individuals with higher educational attainment. Furthermore, these results assert the usefulness of selective immigration policy as an important tool for entrepreneurship development.

2.7 Robustness Checks

Many studies in labor economics either analyse men and women separately or analyse men-only samples due to concerns that women’s labor force participation is interrupted by fertility decision. If labor force participation of women significantly varies across years, in ways not accounted for by the year fixed effects and the state specific linear trends, analysis of pooled samples may induce bias. In Table 2.8, I report separate estimates for samples of men and women. The estimates generated in LPM and TSLS models are consistent with the main results and suggest that the crowding in observed in incorporated self-employment is mainly driven by the crowding in of men.

The preceding analysis models self-employment as a linear function of relevant controls. When estimating marginal effects of a regressor on limited dependent variables, a linear model may yield biased and inconsistent results if a large number of predicted probabilities are outside the unit interval (Horrace & Oaxaca, 2006). Therefore I check if the baseline estimates generated in Linear Probability Model substantially differ from those obtained in Probit models. Table 2.9 presents the results generated by the two model-types side by side. In the case of Probit model, marginal effects are reported and the corresponding standard errors are calculated by delta method. Estimates are similar across model types.
Table 2.8: Impact of Supply of Immigrant Entrepreneurs on Entrepreneurial Propensities of Native Men and Women

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPM</td>
<td>TSLS</td>
</tr>
<tr>
<td>Men</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immigrant entrepreneurs</td>
<td>-0.0003</td>
<td>-0.0022</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>Observations</td>
<td>319,979</td>
<td>319,979</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immigrant entrepreneurs</td>
<td>-0.0004</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>Observations</td>
<td>301,981</td>
<td>301,981</td>
</tr>
</tbody>
</table>

Note: Dependent variable is a binary indicator for whether a native individual is an entrepreneur. The supply of immigrant entrepreneurs is log transformed. All models include control for time varying state characteristics (annual unemployment rate, average annual wage earnings, and population), demographic characteristics (age, marital status, education, race, number of children, city residence status, indicator for public health insurance, and indicator for health insurance coverage through spouse’s policy), state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors adjusted for clustering at state level reported in parentheses. *p < .10. ** < .05. *p < .01

Table 2.9: Comparison of Estimates from LPM and Probit Models

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Unincorporated</th>
<th>Incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPM</td>
<td>Probit</td>
</tr>
<tr>
<td>Immigrant entrepreneurs</td>
<td>-0.0004</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>Observations</td>
<td>621,960</td>
<td>621,960</td>
</tr>
</tbody>
</table>

Note: Dependent variable is a binary indicator for whether a native individual is an entrepreneur. The supply of immigrant entrepreneurs is log transformed. For probit model, average marginal effects are reported and the corresponding standard errors are calculated by delta method. All models include control for time varying state characteristics (annual unemployment rate, average annual wage earnings, and population), demographic characteristics (age, marital status, education, race, number of children, city residence status, indicator for public health insurance, and indicator of health insurance coverage through spouse’s policy), state fixed effects, year fixed effects and state specific linear trends. Robust standard errors adjusted for clustering at the state level reported in parentheses. *p < .10. ** < .05. *p < .01
2.8 Conclusion

The debate on the impact of immigration on the welfare of natives is far from settled. Previous economic research on the topic has mainly centered on the impact of immigration on wages and employment prospects of natives. Although studies frequently mention immigrant entrepreneurship as one of the benefits of immigration, very little work has been done to establish a causal link, if any, between immigrant entrepreneurship and native entrepreneurship. This study has exploited a policy variation across states and years to examine the causal impact of an exogenous change in supply of immigrant entrepreneurs on native entrepreneurial propensity.

At least four important results emerge from the current analysis. The first is that the impact of immigrant entrepreneurs on the entrepreneurial propensities of natives depends on the type of businesses the immigrants operate. Immigrants operating incorporated businesses increase the propensities of native self-employment. However, the immigrants operating unincorporated businesses, on average, have no discernible impact on native self-employment. The significant spillover benefits of incorporated immigrant entrepreneurs are not surprising. Empirical papers in immigration literature frequently find positive effects of high skilled immigrants on labor market outcomes of high skilled natives and negative effects of low-skilled immigrants on the wages and employment prospects of low skilled natives. As incorporated businesses are mostly operated by educated immigrants and the unincorporated businesses mostly by less educated immigrants, the results obtained in this paper are consistent with that finding.

The second important finding of the paper is that the crowding in observed in the incorporated self-employment is driven by crowding in of high skilled natives. This seems to suggest that low skilled natives are not able to capitalize on the benefit that comes with presumably high quality businesses operated by incorporated immigrant
entrepreneurs. However, the low skilled natives may be benefitting through increased employment prospects and higher wages resulting from immigrant-owned businesses. An important extension to this research will be to examine if earnings and employment rates of natives are positively affected by immigrant-owned incorporated businesses.

Third, the sectoral analysis demonstrates that, among those analysed, the trade and transportation sectors experience the strongest crowding in of the native entrepreneurs as a result of immigrant entrepreneurship. The positive impact in trade sector fits well with the finding, in trade literature, that immigrants facilitate trade through knowledge sharing and technology diffusion. The crowding in observed in the transportation sector can be viewed under same light because trade and transportation are closely related sectors. A growing trade sector is likely to cause growth in transportation sector because the output of the later is often the input in the former.

Finally, there is some evidence that immigrants operating unincorporated businesses in manufacturing sector crowd out natives from unincorporated self-employment in that sector. This is possible if immigrant entrepreneurs have access to cheaper inputs, for reasons discussed in Section 3, and are able to cut costs of production. Immigrant entrepreneurs are likely to have advantage over natives especially in the manufacturing sector due to convenience in off shoring.

This study is not without limitations. Due to imperfect data, the sectoral analysis does not unequivocally inform immigration policy regarding which sectors bring in most spillover benefits. More granular data would enable research on specific sectors such as information technology where more recent skilled immigrants are concentrated. Additionally, current analysis is not sufficient to empirically tease out the mechanisms through which immigrant entrepreneurs affect native entrepreneurial propensities. For example, do immigrant entrepreneurs influence the survival rates of the native-owned businesses or their creation rates? Similarly, are immigrants able to cut costs of production or they are accepting lower profits? These questions remain
Nevertheless, this study has provided the first quasi-experimental evidence for the causal link between immigrant entrepreneurship and native entrepreneurship. The results help us fully appreciate the heterogeneity underlying the impact of immigrant entrepreneurs and improve our understanding of the determinants of entrepreneurship in the United States. The findings have implications for policies related to immigration and entrepreneurship development.
CHAPTER 3

CHILDREN’S HEALTH INSURANCE BENEFIT AND LABOR SUPPLY: EVIDENCE FROM NEWLY ARRIVED IMMIGRANTS

3.1 Introduction

A major concern in implementing a means tested public benefit program is its effect on labor supply behaviour of potential beneficiaries. Impending distortions in labor supply, if significant, can alter the overall attractiveness of the benefit program under consideration. The concern has motivated a large number of studies on labor supply impacts of public benefit programs like unemployment insurance, cash transfers to poor families, and publicly funded health insurance coverage for low income families (Cullen & Gruber, 2000; Dave, Decker, Kaestner, & Simon, 2015; Erosa, Fuster, & Kambourov, 2012; Fraker & Moffitt, 1988; Galiani & McEwan, 2013; Yelowitz, 1995). However, the issue of how providing publicly funded health insurance coverage for
children may affect the labor supply of adults has received very little attention.

Since child care costs are important determinants of adult labor supply (Heckman, 1974; D. M. Blau & Robins, 1988; González, 2013), public health insurance for children is likely to play a significant role in adults’ choice of whether to work and how much to work. However, public health insurance coverage for children can affect labor supply of adults in numerous ways making it difficult to sign the impact a priori. On one hand, it can increase families’ disposable income by reducing or eliminating insurance premiums that would otherwise be paid for private coverage. The public coverage can also reduce out-of-pocket expenses related to children’s medical care. This may result in a reduction in labor supply, both at the extensive and at the intensive margin, due to income effect.\(^1\)

On the other hand, public health insurance coverage for children can encourage adult labor force participation by reducing the risk of leaving child at home or at day care. The increase in labor supply, at the intensive margin, may also result from fewer no-shows at work due to better health of children achieved through higher levels of medical access (Kuhlthau & Perrin, 2001).

In this study, I exploit the natural experiment provided by the variation, across states, in immigrants’ eligibility for the State Child Health Insurance Program (SCHIP) to evaluate the impact of a publicly funded children’s health insurance benefit on labor supply of adults. I draw data from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS) and implement a difference-in-difference-in-differences (DDD) strategy that eliminates the effects of a range of confounding variables. I find that children’s health insurance benefit has significant negative effect on labor supply of women, both at the extensive and the intensive margin, but positive effect on labor force participation of men. Point estimates suggest that the SCHIP decreased women’s labor force participation by approximately 7.4

\(^1\)A higher level of unearned income can increase the reservation wage by increasing the value of leisure (Imbens, Rubin, & Sacerdote, 2001), thus influencing the labor force participation.
percentage points and annual hours worked by approximately 15.6%. However, the provision increased labor force participation of men by 5.5 percentage points. I do not find any effect on the annual number of hours worked, conditional on participation, by men.

3.2 Background

3.2.1 Previous Studies

Although the literature on labor supply effects of welfare programs is massive, the question of how providing publicly funded health insurance to children affects the labor supply of adults has been scantily explored. To the best of my knowledge, only two previous studies investigate the effect of public health insurance for children on the labor supply of adults. Perhaps for lack of other sources of identification, both of them exploit SCHIP as a natural experiment.

Tomohara and Lee (2007) perform a difference-in-differences analysis to examine the impact of SCHIP on labor supply of married women. Using data from the ASEC, popularly known as the March CPS, they find no impact on the full sample. H. J. Lee and Tomohara (2008) essentially perform the same exercise, and conclude that SCHIP caused a decline in the labor supply of only particular groups of women, especially the non-white wives, and the married women with pre-school children.

While useful as preliminary assessments, a serious methodological limitation common to both of the above studies is that they define their control and treatment groups on the basis of whether individuals qualify for SCHIP based on the income thresholds set by respective states. Since income is endogenously determined by individual’s choice of labor supply, the treatment in such framework is not exogenous. This makes

---

2A review of theoretical and empirical studies on the labor supply of effects of various welfare programs can be found in Moffitt (2002).
the estimates in those studies prone to bias and inconsistency (Meyer, 1995). The endogeneity is further exacerbated by the measurement errors in respondent reported income (Bollinger, 1998).

In this study, I circumvent the treatment endogeneity problem by defining the treatment and control groups on the basis of SCHIP eligibility rules for immigrants that vary across states and years. The new framework ensures the conditional ignorability of the treatment status required for the unbiased estimation of the average treatment effect (ATE).\(^3\)

### 3.2.2 SCHIP and Immigrants

An important step of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 was to ban the immigrants who have been in the US for less than five years, henceforth new immigrants, from receiving any federal means tested benefits including the Medicaid.\(^4\)

After the US congress passed State Child Health Insurance Program (SCHIP) to cover uninsured children in 1997 all states implemented some form of SCHIP by 2000. Taking advantage of the ample autonomy the states had in designing the SCHIP, 15 of them including the District of Columbia (DC), henceforth generous states, used State-only funds to include the children of new immigrants in their SCHIP. The other 36 states, referred to as less generous states hereafter, did not. This essentially created a natural experiment whereby new immigrants in the generous states were treated with public health insurance for children (hence the treatment group) and those in less generous states were not (the control group). This categorization of control and treatment group is superior to defining the groups on the basis of income eligibility

---

\(^3\)A conditionally ignorable treatment simply means that the treatment status is independent of the outcome given the covariates. In other words, \((Y_{T=1}, Y_{T=0}) \perp T \mid X\) for a binary treatment variable \(T\), outcomes \(Y\), and covariates \(X\).

\(^4\)Means tested benefits are those for which individuals and families qualify only if their past year income is below a certain threshold.
because immigrants cannot endogenously choose a ‘new immigrant’ status.\(^5\)

3.3 Data and Methods

3.3.1 Data


In addition to the detailed demographic information, including the immigrant status and the year of entry into the country, the ASEC provides comprehensive information on the employment status, occupation, industry, number of weeks worked in a year, and usual hours worked per week. The universe includes all 25-64 year old immigrant adults residing in families with at least one child.

There are two outcome variables of interest in this analysis. One is a binary indicator denoting the labor force participation status and the other is a continuous variable measuring number of hours worked in a year\(^7\).

Given the rich data set, I am able to control for a range of demographic variables relevant for labor supply decisions. Specifically I control for age, education, marital status, household size, race, and city resident status. Additionally, the repeated cross-

---

\(^5\)To the extent that some new immigrants might be able to make themselves eligible for SCHIP by finding ways to naturalize, or changing states of residence, this assumption can be violated. I address the first concern by excluding all naturalized new immigrants from the analysis. To see if immigrants are crossing state borders in response to SCHIP generosity, I examine the pattern of immigrant arrival to generous states before and after SCHIP implementation. No distinct change in trend is observable. Furthermore, there is very little evidence that inter-state movement of immigrants is motivated by availability of welfare benefits (Frey, Liaw, Xie, & Carlson, 1996; P. B. Levine & Zimmerman, 1999).

\(^6\)Data preceding 1994 cannot be used because no question on immigrant status was asked. In 2009, the Child Health Insurance Program Reauthorization Act (CHIPRA) allowed all states to cover immigrant children, irrespective of when they entered the country, with federal funds. Thus the source of identification persists only until 2009.

\(^7\)All individuals reporting to be not working and not looking for work are classified as out of labor force. I construct the ‘number of hours worked in a year’ variable by multiplying the number of hours worked per week and the numbers of weeks worked in year.
sectional nature of data allows me to control for state fixed effects, year fixed effects and state-specific linear trends to account for macro-economic differences across states and years. The summary statistics on relevant variables are presented in Table 3.1.

### 3.3.2 Methods

The starting point in this analysis is to estimate a simple a difference-in-differences (DD) model to compare the labor supply decisions of newly arrived immigrants in generous states versus less generous states. This can be accomplished by estimating the model below:

\[
Y_{ist} = \alpha + \beta_1.Gen_{s}.Post_{st} + \beta_2.Gen_{s} + \beta_3.Post_{st} + X'_{ist} \cdot \gamma \\
+ \eta_s + \lambda_t + \theta.t\eta_s + \epsilon_{ist} \quad (3.1)
\]

where \(Y_{ist}\) is the particular outcome variable (i.e. either labor force participation or the total number of hours worked in a year) associated with immigrant \(i\), state \(s\) and year \(t\), \(Gen_s\) indicates whether the state is a generous state, \(Post_{st}\) indicates whether the year is after the SCHIP implementation in the state \(s\), \(X'_{ist}\) is a vector of individual characteristics potentially correlated with the outcome variables, and \(\epsilon_{ist}\) is random error term. The \(\beta\)s and \(\gamma\) are the parameters of the model. \(\eta_s\), \(\lambda_t\), and the term \(\theta.t\eta_s\) respectively capture the state fixed effects, year fixed effects, and state specific linear trends. When the usual assumptions of a classical linear regression model are met, the parameter \(\beta_1\) gives the effect of providing SCHIP benefits to newly arrived immigrants on their labor supply.

An important limitation of the ordinary DD model is that it doesn’t account for the possibility that there may be non-SCHIP factors that differentially affect the labor supply decisions of individuals in generous and less generous states. One way
<table>
<thead>
<tr>
<th>Variable</th>
<th>Women</th>
<th></th>
<th>Men</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observations</td>
<td>Mean</td>
<td>Standard</td>
<td>Observations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>deviation</td>
<td></td>
</tr>
<tr>
<td>In labor force</td>
<td>66726</td>
<td>.635</td>
<td>.481</td>
<td>58641</td>
</tr>
<tr>
<td>New<em>Gen</em>Post</td>
<td>66726</td>
<td>.058</td>
<td>.234</td>
<td>58641</td>
</tr>
<tr>
<td>New*Post</td>
<td>66726</td>
<td>.119</td>
<td>.324</td>
<td>58641</td>
</tr>
<tr>
<td>New*Gen</td>
<td>66726</td>
<td>.085</td>
<td>.280</td>
<td>58641</td>
</tr>
<tr>
<td>Post*Gen</td>
<td>66726</td>
<td>.430</td>
<td>.495</td>
<td>58641</td>
</tr>
<tr>
<td>New</td>
<td>66726</td>
<td>.162</td>
<td>.369</td>
<td>58641</td>
</tr>
<tr>
<td>Gen</td>
<td>66726</td>
<td>.576</td>
<td>.494</td>
<td>58641</td>
</tr>
<tr>
<td>Post</td>
<td>66726</td>
<td>.780</td>
<td>.414</td>
<td>58641</td>
</tr>
<tr>
<td>Age</td>
<td>66726</td>
<td>38.523</td>
<td>8.366</td>
<td>58641</td>
</tr>
<tr>
<td>Education</td>
<td>66726</td>
<td>2.320</td>
<td>1.267</td>
<td>58641</td>
</tr>
<tr>
<td>Household size</td>
<td>66726</td>
<td>4.723</td>
<td>1.757</td>
<td>58641</td>
</tr>
<tr>
<td>Race</td>
<td>66726</td>
<td>2.832</td>
<td>.928</td>
<td>58641</td>
</tr>
<tr>
<td>City resident</td>
<td>66455</td>
<td>.932</td>
<td>.250</td>
<td>58397</td>
</tr>
<tr>
<td>Married</td>
<td>66726</td>
<td>.778</td>
<td>.415</td>
<td>58641</td>
</tr>
<tr>
<td>Annual hours</td>
<td>41372</td>
<td>1714.783</td>
<td>703.4543</td>
<td>54702</td>
</tr>
</tbody>
</table>
to account for such possibility is to get a difference between the labor supply of ‘old’ immigrants, in generous versus the less generous states, that are not affected by the policy change and subtract the difference from DD estimate. This can be achieved by estimating a difference-in-difference-in-differences (DDD) model.

The triple-difference identification scheme allows for differences between generous and less generous states, differences between individuals over time, and differences between immigrants who are new and those who are not. The regression takes the form:

\[
Y_{ist} = \alpha + \beta_1 \cdot \text{New}_i \cdot \text{Gen}_s \cdot \text{Post}_{st} + \beta_2 \cdot \text{New}_i \cdot \text{Gen}_s + \beta_3 \cdot \text{New}_i \cdot \text{Post}_{st} + \beta_4 \cdot \text{Gen}_s \cdot \text{Post}_{st} + \beta_5 \cdot \text{New}_i + \beta_6 \cdot \text{Gen}_s + \beta_7 \cdot \text{Post}_{st} + \mathbf{X}_{ist}' \cdot \gamma \\
+ \eta_s + \lambda_t + \theta \cdot t \eta_s + \epsilon_{ist} \tag{3.2}
\]

where, as mentioned earlier, \( Y_{ist} \) is the particular outcome variable (i.e. either labor force participation or the total number of hours worked in a year) associated with immigrant \( i \), state \( s \) and year \( t \), \( \text{New}_i \) indicates whether the immigrant \( i \) is a new immigrant, \( \text{Gen}_s \) indicates whether the state is a generous state, \( \text{Post}_{st} \) indicates whether the year is after the SCHIP implementation in the state \( s \), \( \mathbf{X}_{ist}' \) is a vector of individual characteristics potentially correlated with the outcome variables, and \( \epsilon_{ist} \) is random error term. The \( \beta_s \) and \( \gamma \) are the parameter of the model. \( \eta_s \), \( \lambda_t \), and the term \( \theta \cdot t \eta_s \) respectively capture the state fixed effects, year fixed effects, and state specific linear trends.

In the model represented by Equation 3.2, \( \beta_1 \) gives the impact of the SCHIP on labor supply behaviour of new immigrants. Intuitively, the coefficient represents the change in outcome variable during the post-SCHIP period among new immigrants who reside in generous states, net of any changes in ‘old’ immigrants, and any under-
lying differences between the states. The triple difference strategy is superior to basic difference-in-differences because the former subtracts out the trends that may differentially affect the treatment and the control groups, making the ‘common trends’ assumption unnecessary.

Since the labor force participation rates of men and women are significantly different from each other (F. D. Blau & Kahn, 2007), I separately estimate regressions for men and women.

### 3.4 Results and Discussion

First I examine how the treatment group differs from the control group in the pre and post treatment period. Figure 3.1 plots the labor force participation rate and the annual number of hours worked by new immigrant adults in generous and less generous states at different points of time before and after SCHIP implementation. Two important observations can be made from graphical representation in Figure 3.1. First, the pre-treatment trends in labor force participation and annual hours worked are fairly comparable in both groups of states making the DD identification strategy a potentially suitable choice. Second, both measures of labor supply show an increasing trend in both groups of states before the treatment but the measures decline right after the treatment in generous states. This preliminary evidence is suggestive of the negative impact of SCHIP on labor supply behavior of new immigrants in generous states.

Table 3.2 presents results generated in the baseline DD model. The first two columns show that provision of SCHIP decreased labor force participation of women by 8.7 percentage points but increased that of men by 5.1 percentage points. The third and the fourth columns in the table show that the implementation of SCHIP did not have any effect on the annual hours worked. As stated earlier, these estimates
Figure 3.1: Pre and Post SCHIP Labor Supply
Table 3.2: Impact of SCHIP on Labor Supply of Men and Women: DD Estimates

<table>
<thead>
<tr>
<th></th>
<th>Labor force Participation</th>
<th>Log annual hours worked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td>( \beta_{\text{Gen}\times\text{Post}} )</td>
<td>-0.087** (0.040)</td>
<td>0.051** (0.021)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0594</td>
<td>0.0569</td>
</tr>
<tr>
<td>Observations</td>
<td>10,745</td>
<td>7,934</td>
</tr>
</tbody>
</table>

Note: Coefficients are estimates from DD regression including demographic controls, state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors clustered at the state level reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% level.

are prone to bias and inconsistency due to violation of common trends assumption.

Next I present the results from the preferred DDD model. Table 3.3 presents the DDD estimates of the impact of SCHIP on labor force participation rate and annual hours worked. The coefficients indicate negative impact of SCHIP on both the labor force participation rate and the average annual hours worked of women. Specifically, the SCHIP reduced the labor force participation of women by 8.7 percentage points and annual hours worked by 15.6%. The DDD estimates also show that the labor force participation of men increased by 5.5 percentage points. There is no statistically significant effect on the annual number of hours worked by men.

Negative impact on labor supply of women, but positive on that of men, at least at the extensive margin, is consistent with traditional roles of men and women in families. It is possible that increased disposable income, due to publicly funded health insurance for children, induced some women to withdraw from labor force and/or reduce the hours worked. Although increase in participation of men as a result of SCHIP is slightly puzzling, one possibility is that availability of SCHIP produced an environment more conducive for child bearing, thus encouraging women
Table 3.3: Impact of SCHIP on Labor Supply of Men and Women: DDD Estimates

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{\text{New\ Gen\ Post}}$</td>
<td>-0.074***</td>
<td>0.055**</td>
<td>-0.156**</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.041)</td>
<td>(0.058)</td>
<td>(0.706)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0792</td>
<td>0.0426</td>
<td>0.0325</td>
<td>0.0338</td>
</tr>
<tr>
<td>Observations</td>
<td>66,455</td>
<td>58,397</td>
<td>41,203</td>
<td>54,470</td>
</tr>
</tbody>
</table>

Note: Coefficients are estimates from DDD regression including demographic controls, state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors clustered at the state level reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% level.

to withdraw for fertility reasons and men to increase participation to compensate losses in income due to women’s withdrawal from labor force or reduction in hours. However, this remains only a conjecture at this point.

### 3.5 Robustness Checks

The results obtained in models estimated so far are obtained by comparing behaviour of individuals in households with at least one child. It is reasonable to argue that child bearing itself is an endogenous decision and some households might be incentivized to have children as a result of the SCHIP provision. If that was the case, then the DDD estimates obtained in above models would not necessarily be the ATE for whole immigrant population. One way to get closer to estimating an ATE for whole population is to include in the analysis newly arrived individuals from all households, irrespective of whether the households have any children. The results generated in these models are presented in Table 3.4. The results are qualitatively and quantitatively similar to the results presented in Table 3.3.
Table 3.4: Impact of SCHIP on Labor Supply of Men and Women: DDD Estimates with All Immigrant Households

<table>
<thead>
<tr>
<th></th>
<th>Labor force Participation</th>
<th>Annual hours worked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td>$\beta_{\text{NewGenPost}}$</td>
<td>-0.043**</td>
<td>0.049***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0822</td>
<td>0.0440</td>
</tr>
<tr>
<td>Observations</td>
<td>109,783</td>
<td>104,767</td>
</tr>
</tbody>
</table>

Note: Coefficients are estimates from DDD regression including demographic controls, state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors clustered at the state level reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% level.

Another set of robustness check involves comparing weighted and unweighted estimates. Often a large difference between coefficients estimated in weighted and unweighted models can be a warning sign of specification error. Because all the results obtained so far are from weighted least squares models, I estimate the unweighted DDD regressions and present results in Table 3.5. Estimates from unweighted models are fairly close to those from weighted models indicating reasonably correct specifications.

One concern in estimating the treatment effect by difference-in-difference method is that the estimator may simply be picking up the effect of other changes taking place around the policy implementation time. In other words, the policy variation may just be a coincidence and the effects given my DDD estimator are wrongly attributed to the policy change. A set of falsification tests, known as placebo tests, can be conducted to see if falsely assigning policy changes gets us the same effects obtained in main results. If placebo treatments do not show significant effects as observed in the main results, it is an indication that the effects given by DDD estimator are indeed due to treatment.
Table 3.5: Impact of SCHIP on Labor Supply of Men and Women: Unweighted Estimates

<table>
<thead>
<tr>
<th>β_{New<em>Gen</em>Post}</th>
<th>Women</th>
<th>Men</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>β_{New^*Gen^*Post}</td>
<td>-0.074***</td>
<td>0.040*</td>
<td>-0.215***</td>
<td>-0.023</td>
</tr>
<tr>
<td>(0.023)</td>
<td>(0.022)</td>
<td>(0.057)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0781</td>
<td>0.0361</td>
<td>0.0314</td>
<td>0.0320</td>
</tr>
<tr>
<td>Observations</td>
<td>66,455</td>
<td>58,697</td>
<td>41,203</td>
<td>54,470</td>
</tr>
</tbody>
</table>

Note: Coefficients are estimates from DDD regression including demographic controls, state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors clustered at the state level reported in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% level.

In Table 3.6, I present absolute values of t-statistic associated with the coefficients of interest when placebo treatments are assigned one period, two period and three period ahead of actual treatment. The table shows that the size of t-statistics are progressively smaller as the placebo treatment occurs farther from the actual implementation. This is an indication that the effects given by DDD estimator are plausibly the effects of the variation in SCHIP eligibility rules.

However, there are two anomalies to take note of. First, the placebo treatments one period and two periods before the actual policy implementation show significant effect on women’s labor force participation. This can happen if individuals change their labor supply decision in anticipation of the treatment, which seems likely in the case at hand. Although SCHIP was passed by the federal government in 1997, the state level implementations happened over years through 2000. The second anomaly in the falsification test is that the placebo treatment one period before the actual policy implementation shows significant effect on annual hours worked of men. This is somewhat puzzling but this seems to have happened merely by chance because all other placebos and actual treatment show no significant effect on annual hours.
Table 3.6: Falsification Tests with Placebo Treatments

<table>
<thead>
<tr>
<th>Timing</th>
<th>Labor force Participation</th>
<th>Annual hours worked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Women</td>
<td>Men</td>
</tr>
<tr>
<td>Post policy</td>
<td>2.71**</td>
<td>2.10**</td>
</tr>
<tr>
<td>One period before policy</td>
<td>2.38**</td>
<td>0.48</td>
</tr>
<tr>
<td>Two periods before policy</td>
<td>1.99**</td>
<td>0.79</td>
</tr>
<tr>
<td>Three periods before policy</td>
<td>0.88</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*Note: Reported numbers are the absolute values of t-statistics from DDD regression including demographic controls, state fixed effects, year fixed effects, and state specific linear trends. ***, **, and * indicate significance at 1%, 5%, and 10% level.*

3.6 Conclusion

Exploiting the policy change and the data previously used in the literature, but making use of a novel source of identification, this study examines the impact of publicly funded children’s health insurance benefit on labor supply of adults. The study finds significant negative impact of publicly funded health insurance benefit for children on the labor supply of women, both at the extensive as well as intensive margins. The benefit, however, has positive effect on the labor force participation of men.

Since the results are obtained by examining the labor supply behaviour of newly arrived immigrants, the generalizability of these results depends on the extent to which labor supply behaviour of new immigrants is comparable to the rest of the population. Because newly arrived immigrants are significantly different from the rest of the population in terms of social capital, economic well-being, educational achievement, etc. it is possible that the effects of the benefit observed in this population subgroup are different from other sub groups. Nevertheless, the results underscore
the labor supply distortions associated with welfare benefits and thus highlight the importance of optimal mechanism design in welfare economics.
References


   Journal of Labor Economics, 21(3).
icy, 5(3), 160–188.


Hicks, J. R. (1932). The theory of wages.


Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost inno-


(Available at [http://www.hbs.edu/faculty/Publication7-be52477123ee58d0.pdf](http://www.hbs.edu/faculty/Publication7-be52477123ee58d0.pdf))


Ross, D. C., Cox, L., & Marks, C. (2007). Resuming the path to health coverage for


Appendices
Appendix A

A.1 Generous SCHIP, Health Insurance Coverage of Native Children and Eligibility Guidelines

I estimate the following regression to test the hypothesis:

\[
\text{Insurance}_{ist} = \beta_0 + \beta_1 . Gen_s . Post_{st} + \beta_2 . Gen_s + \beta_3 . Post_{st} + \eta_s + \lambda_t + \gamma . t \nu_s + \epsilon_{ist} \quad (A.1)
\]

where \( \text{Insurance} \) is an indicator for whether a native child has insurance coverage, \( Gen \) and \( Post \) are indicators for SCHIP generosity in a state and whether the year is after SCHIP implementation respectively; \( \eta_s \), \( \lambda_t \), and \( \gamma . t \nu_s \) are state fixed effects, year fixed effects, and state specific linear trend respectively; and \( \epsilon \) is error term. \( \beta_1 \) gives the impact of generous SCHIP on the likelihood of a native child having an insurance coverage.
A.2 Industry Classification in CEPR Extracts

Table A.1: Impact of Generous SCHIP on Native Children’s Health insurance Coverage

<table>
<thead>
<tr>
<th>Dependent variable: Any health insurance coverage</th>
<th>Medicaid or SCHIP coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen.Post</td>
<td>-0.0123</td>
</tr>
<tr>
<td></td>
<td>(0.0102)</td>
</tr>
<tr>
<td>Observations</td>
<td>416,761</td>
</tr>
</tbody>
</table>

Note: Dependent variable is whether a native child has a health insurance coverage. Coefficients represent DD estimates from Ordinary Least Squares regression with state fixed effects, year fixed effects, and state specific linear trends. Robust standard errors adjusted for clustering at state level are reported in parentheses. *p<.10. **p<.05. ***p<.01
Figure A.1: State Eligibility Guidelines for Children’s Medicaid/SCHIP

Table A.2: Industry Classification Available in CEPR’s Uniform March Supplement Extracts

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>Agriculture, Forestry, Fishing, and Hunting</td>
</tr>
<tr>
<td>Mining</td>
<td>Mining</td>
</tr>
<tr>
<td>Construction</td>
<td>Construction</td>
</tr>
<tr>
<td>Manufacturing(durable/nondurable)†</td>
<td>Manufacturing(durable/nondurable)</td>
</tr>
<tr>
<td>Transportation</td>
<td>Wholesale Trade</td>
</tr>
<tr>
<td>Communications</td>
<td>Retail Trade</td>
</tr>
<tr>
<td>Utilities and Sanitary Services</td>
<td>Transportation and Warehousing</td>
</tr>
<tr>
<td>Wholesale Trade</td>
<td>Utilities</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>Information</td>
</tr>
<tr>
<td>Banking and Other Finance</td>
<td>Finance and Insurance</td>
</tr>
<tr>
<td>Insurance and Other Finance</td>
<td>Real Estate and Rental and Leasing</td>
</tr>
<tr>
<td>Private Household Services</td>
<td>Professional, Scientific, and Technical services</td>
</tr>
<tr>
<td>Business Services</td>
<td>Management, Administrative Support and Waste Management Services</td>
</tr>
<tr>
<td>Repair Services</td>
<td>Educational services</td>
</tr>
<tr>
<td>Personal Services, Except Private Household</td>
<td>Health Care and Social assistance</td>
</tr>
<tr>
<td>Entertainment and Recreation Services</td>
<td>Art, Entertainment and Recreation</td>
</tr>
<tr>
<td>Hospitals</td>
<td>Accommodations and Food Services</td>
</tr>
<tr>
<td>Health Services, Except Hospitals</td>
<td>Private Households</td>
</tr>
<tr>
<td>Educational Services</td>
<td>Other Services</td>
</tr>
<tr>
<td>Social Services</td>
<td>Public Administration</td>
</tr>
<tr>
<td>Other Professional Services</td>
<td>Armed Forces and Active Military</td>
</tr>
<tr>
<td>Forestry and Fisheries</td>
<td>Never Worked</td>
</tr>
<tr>
<td>Public Administration</td>
<td>Never Worked/Last Job Armed Forces</td>
</tr>
</tbody>
</table>

Note: †Further breakdown available.