2022

Essays on Immigration Policy

Andres Felipe Mira
University of Kentucky, afmirav@gmail.com
Digital Object Identifier: https://doi.org/10.13023/etd.2022.311

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Andres Felipe Mira, Student

Dr. Christopher Bollinger, Major Professor

Dr. Carlos Lamarche, Director of Graduate Studies

ESSAYS ON IMMIGRATION POLICY

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy in the
Gatton College of Business and
Economics at the University of Kentucky

By
Andrés F. Mira
Lexington, Kentucky

Co-Directors: Dr. Christopher Bollinger, Professor of Economics
and Dr. Aaron Yelowitz, Professor of Economics
Lexington, Kentucky
2022

ABSTRACT OF DISSERTATION

ESSAYS ON IMMIGRATION POLICY

Considerable interest has been placed on the subject of what to do with the sizable undocumented population currently residing in the United States. Having a full understanding of the size of this population and the impact immigration policy has on them is of critical importance to policy makers. Data limitations in nationally-representative surveys have limited the analyzes of the effects of immigration policy in the academic literature. As such, this dissertation consists of three essays contributing to the literature on the impact of immigration policy and on identifying unobserved populations.

In my first essay, I examine the labor market response of undocumented youth that participated in the Deferred Action for Childhood Arrivals (DACA) program which provides them temporary deportation relief and work authorization. I use data from the U.S. Citizenship and Immigration Services to construct a probabilistic measure for unobserved DACA participation. Using the American Community Survey (ACS), I estimate a two-sample model of the effect of participating in the DACA program. I also estimate spillover effects of DACA on eligible but non-participating undocumented youth. I find that DACA significantly improved labor market and education outcomes of DACA recipients, with magnitude of the treatment-on-the-treated effects at least twice as large as the intent-to-treat estimates obtained from using only the observed eligibility indicator typically used in literature. Evidence of a negative spillover effect on eligible non-participants is documented with a decrease in labor force participation and school attendance.

My second essay considers nonsampling error due to item nonresponse in the estimates of the size and legal composition of the foreign-born population produced using the ACS. The standard practice to address item nonresponse is to impute values under the assumption that nonresponse is conditionally random. I form credible interval estimates that make no assumptions about the values of missing data by considering all uncertainty due to item nonresponse. Without this assumption, the size of the foreign-born population in the US falls somewhere between 40.4 and 59.4 million as of 2019 compared to the Census estimate of 44.9 million. Bounding estimates of the size of the undocumented population fall between

7.3 and 23.3 million compared to the widely accepted estimate of 11 million undocumented immigrants.

In my third essay, I return to analyzing the effects of the DACA program. I examine misclassification bias arising from item-nonresponse in the estimated intent-to-treat effects of the DACA program. Assigning DACA eligibility is based on the responses to specific demographic questions, any of which the individual may not respond to. If the assumption that nonresponse to these questions are conditionally missing fails, this can lead to traditional misclassification bias and attenuate the results when using imputed values. Adjusting for potential misclassification bias by removing non-respondents leads to estimates of the intent-to-treat effects of DACA on labor market outcomes that are 22% to 77% higher than when including non-respondents, depending on the outcome of interest.

KEYWORDS: Immigration, Immigration Policy, DACA, Item-nonresponse, Misclassification Bias

Andrés F. Mira

August 5, 2022

ESSAYS ON IMMIGRATION POLICY

By
Andrés F. Mira

Dr. Christopher Bollinger
Co-Director of Dissertation

Dr. Aaron Yelowitz
Co-Director of Dissertation

Dr. Carlos Lamarche
Director of Graduate Studies

August 5, 2022
Date

DEDICATION

To my mother, Consuelo Vallejo, and my brother, Juan Camilo Mira.

ACKNOWLEDGMENTS

I would first like to thank my advisors Dr. Chris Bollinger and Dr. Aaron Yelowitz. They provided invaluable guidance and support throughout my academic studies. I am appreciative for Aaron's persistent motivation to always strive for the better. I am deeply grateful to Chris's insightful and engaging discussions during our weekly meetings and for his patience at every stage; he is the epitome of the ideal mentor. I would also like to thank my committee members and outside reader respectively: Dr. Bill Hoyt, Dr. Rajeev Darolia, and Dr. Michael Samers. Each provided valuable expertise which improved this work. I would also like to make a special thanks to my colleague, coauthor, and friend, Hyein Kang. I could not have asked for a better friend to go through my doctoral studies with. Finally, I cannot begin to express my gratitude to my mother, Consuelo Vallejo. It would have been impossible for me to complete my Ph.D without her unwavering support and encouragement. I also thank my brother and best friend, Juan Camilo Mira, who has been a role model to aspire to and has made me the person I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

The United States has the largest share of immigrants of any country in the world, of which, a considerable portion are estimated to have no formal legal status. There is intense political debate on whether or not to provide a formal pathway to citizenship to these undocumented immigrants. The ability to accurately estimate the size of the undocumented population and the effects of immigration policy on this population are paramount for policy markers ability to make informed policy decisions.

The vast majority of research on immigration is conducted through the use of survey data. While commonly used nationally-representative datasets have allowed researchers to answer critical questions there are limitations that needs to be noted. There are two main issues when using survey data. The first issue to be considered is simply the lack of data. For instance, most datasets do not include information on an individual's legal status. This limits what type of questions are able to be answered. The second issue regards the quality of the data. Surveys are impacted by both sampling and nonsampling errors such as sample size, nonresponse bias, and misreporting. The quality of the data will limit the accuracy of the estimates produced. This dissertation consists of three essays on immigration policy addressing these concerns directly.

In essay 1 (Chapter 2) and essay 3 (Chapter 4) I focus on the Deferred Action for Childhood Arrivals (DACA) program. Enacted in 2012 through executive memorandum, DACA provided eligible undocumented youth temporary deportation relief and work authorization. This program eliminates the severe labor market frictions caused by having no formal legal status. As of 2018, over 824,000 undocumented youth participated in the program and received these benefits. Of the 824,000 total participants, roughly 680,000 maintain an active status.

Data limitations in nationally-representative surveys have limited the analyses of the effects of DACA in the academic literature. The academic literature on the effects of DACA has focused on estimating the intent-to-treat (ITT) effects, that is, the effects of being eligible for the program on eligible undocumented immigrants. Understanding the ITT effects are of importance as it provides information on the effects of the parameters policy makers can change when formulating policy. The ITT effects are, however, only one side of the coin when evaluating the effects of this policy. Estimating the effects of DACA on those who participated in the program is of similar importance. Particularly in understanding the full costs and benefits of the program and to inform the ongoing discussion on whether to terminate the program and eliminate the benefits currently being enjoyed by 680,000 active recipients.

In Chapter 2, I estimate the effect of participating in DACA on labor market outcomes. The most commonly used survey to study the effects of DACA is the American Community Survey (ACS). While the ACS provides information on a wide range of demographic characteristics on each individual to identify their DACA eligiblity status, there is no information on whether an individual had actually participated in DACA. I address this data limitation by supplementing the ACS with publicly available administrative data from the U.S Citizenship and Immigration Service (USCIS). The USCIS provides information on the total number of DACA recipients by year and country-of-origin. With this information and estimates on the size of the DACA-eligible population, I construct a probabilistic measure for unobserved DACA participation. This allows for the first estimation of the treatment-on-the-treated effects of the DACA program. This method also allows me to estimate the spillover effect on DACA-eligible non-participants. I find that the reduction of labor market frictions through deferred action and work authorization from participating in the DACA program led to considerable labor market benefits and improvement on educational outcomes. I also find eligible non-participating non-citizens experienced negative spillover effects as they withdrew from both schooling and the labor market following the enactment of DACA.

Essay 2 (Chapter 3) diverges from the topic of DACA and focuses on a much broader question: How many immigrants are there in the United states? The most widely accepted estimates of the size and legal composition are derived using survey data in whole or in part. As such, these estimates are impacted by the quality of the survey data and the core assumptions used to adjust for the quality of the data. One critical assumption is that item non-respondents are similar to respondents conditional on a set of observable characteristics. This is typically referred to as the missing at random assumption (MAR). The MAR assumption is a strong assumption in this case as it implies that the distribution of legal status among non-respondents is the same as that of respondents.

I relax the strong, and untested assumption, that nonresponse is missing at random, and I produce interval estimates of the size and legal composition of the US foreign-born population that make no such assumption. These bounds capture the total uncertainty caused by item nonresponse. Taking into account uncertainty from item nonresponse leads to significantly large interval estimates. Bounding estimates of the size the foreign-born population in the US falls somewhere between 40.4 and 59.4 million as of 2019 compared to the Census estimate of 44.9 million. Without the MAR assumption, the size of of the undocumented population fall between 7.3 and 23.3 million compared to the widely accepted estimate of 11 million undocumented immigrants.

Imposing the MAR assumption on the data has broader implications on the economics on immigration literature besides the estimates of the size and legal composition of the foreign-born population. The literature has used these population estimates and imputed values to study the effects of immigration and of immigration policy. Accepting the MAR assumption abstracts away the uncertainty caused by item nonresponse. If the MAR assumption fails, though, these estimated effects will be biased.

With this implication, Chapter 4 returns to the estimated effects of DACA eligiblity on labor market outcomes. The literature on DACA has implicitly accepted the MAR assumption by including item non-respondents in their samples. Identifying an individuals DACA eligiblity requires the response to a number of demographic questions any of which the individual may not have responded to. If the MAR assumption fails for any one of the questions, nonresponse will lead to classical misclassification bias and attenuate the effects of DACA. I perform multiple methods to adjust for potential misclassification bias. Adjusting for this bias by removing non-respondents leads to estimates of the ITT effects of DACA that are 22% to 77% higher than when including non-respondents, depending on the outcome of interest. I also produce bounded estimates using a similar procedure used in Chapter 3 to take into account uncertainty caused by item nonresponse. Interestingly, the bounds fall directly within the estimates when using the whole sample and when using the respondent only sample.

While data limitations create obstacles in analysing immigration policy, alternative methodology can be used to alleviate these restrictions. With quality of survey data diminishing, it is critical to supplement existing datasets with additional data from other surveys or administrative records. It is also important for research to acknowledge the data limitations and adjust for them as it can have a considerable impact in the quality of the estimates produced.

Chapter 2 Estimating the Average Treatment-on-the-Treated Effects of the DACA Program

2.1 Introduction

As of 2017, an estimated 10.5 million undocumented immigrants resided in the United States, representing 3.2% of the total population (Passel and Cohn, 2019). Whether congress should pass amnesty and in what form has been subject to heated debate. Considerable interest has been focused on the most recent attempt to regularize this population, the Deferred Action for Childhood Arrivals (DACA) program. Enacted on June 15, 2012 by President Obama through an executive memorandum, DACA grants eligible undocumented youth temporary protection from deportation and provides work authorization. These benefits are subject to renewal every two years. By doing so, the DACA program eliminates considerable labor market barriers that recipients previously faced due to their lack of legal status.

In this paper, I examine how DACA *participation* affected undocumented youth's labor market behavior. In addition, I test for possible spillover effects of the enactment of the DACA program on eligible non-participants. As of 2018, 824,000 individuals have participated in the DACA program with 680,000 maintaining an active status. Understanding the effect of *participating* in the program is of valuable information to policy makers when evaluating whether to continue or terminate DACA. Understanding the behavioral effects of both participants and of eligible but non-participating undocumented youth will also provide information on the effects of similar proposed amnesty programs considered by congress.

A growing literature has developed estimating the effects of DACA eligibility on various outcomes. Previous work has shown that DACA improves labor market outcomes (Pope, 2016; Amuedo-Dorantes and Antman, 2017), improves health outcomes among children and adults (Venkataramani et al., 2017; Hainmueller et al., 2017; Giuntella and Lonsky, 2020), reduces teenage pregnancy (Kuka et al., 2019), reduces the propensity to commit crime (Gunadi, 2019), and improves their sleep behavior (Giuntella et al., 2021). The effect of DACA eligibility on education has been mixed. Pope (2016), Amuedo-Dorantes and Antman (2017), and Hsin and Ortega (2018) show negative or insignificant effects on school attendance for those that are immediately eligible (meet the education requirement) while Kuka et al. (2020) and Ballis et al. (2020) find positive effects among those that may want to become eligible but do not yet meet the education requirement. Ballis et al. (2020) is the only other paper to estimate spillover effects of the DACA program. Ballis et al. (2020) finds evidence of spillover effects in educational achievement among students with more DACA-eligible peers.

The previous studies have focused exclusively on the effects of DACA on those that are eligible for the program. This has limited the discussion to the average intent-to-treat effects of DACA. The ability to extrapolate the average treatment-on-the-treated effects from the average intent-to-treat effects estimated in the literature requires the following strong assumptions: (1) no self-selection into the program and (2) no spillover effects on the non-participating population. Hipsman et al. (2016) estimate that 37% of immediately eligible undocumented immigrants did not participate in DACA. Given that these are strong assumptions the treatment-on-the-treated effects need to be estimated directly. Additionally, the assumption that eligible non-participants are not affected indirectly through increased labor market competition or behavioral changes from the changing legal environment needs to be empirically tested.

Researchers have also been limited to using lack of citizenship along with ethnicity as a proxy for undocumented status as there is no indicator for undocumented status in large representative datasets. With an estimated 39% of non-citizens age 18 to 35 being documented immigrants (Baker and Rytina, 2013), the current observable measure of eligibility are significantly contaminated with authorized non-citizens. Having up to 39% of the observed DACA-eligible population being false-positives will lead to estimates of the average intent-to-treat effects of DACA to be considerably attenuated towards zero. Only one other paper, Ballis et al. (2020), uses administrative data to reduce measurement error in the eligiblity indicator. Ballis et al. (2020) uses variation in the number of DACA applications relative to the size of the foreign-born population across zip code in Los Angeles in 2014 to proxy for likely-undocumented status. Focusing on only Hispanic or Mexican non-citizens also misses important heterogeneous effects across nationality and ethnicity. Understanding how non-Hispanics are affected by conditional amnesty is becoming more pertinent for policy makers as this group continues to become a greater share of the undocumented population (Passel and Cohn, 2019).

In this paper, I expand on this literature by using publicly available administrative data from the U.S citizenship and Immigration Services (USCIS) to construct a probability measure of DACA participation. The USCIS data provides a total count of DACA recipients by country-of-origin for each year since the enactment of DACA up to 2018. Combining estimates of the DACA-eligible non-citizen population by country-of-origin using the ACS with the USCIS data allows me to construct a probabilistic measure for the unobserved DACA participation across country-of-origin and time. To estimate the treatment-on-the-treated effects of DACA on those who participated in the program, I merge this measure into the American Community Survey (ACS) and estimate a two-sample model. These are the best estimates available of the treatment-on-the-treated effects of DACA with currently available administrative data. In an alternative model, I include a probabilistic measure on the likelihood an eligible non-citizen is not a DACA recipient along with the probabilistic measure. This allows me to estimate both the spillover effects of DACA on eligible non-participants along with the direct treatment effects on DACA recipients using ineligible non-citizens as the control group.

I find that DACA significantly improved the labor market outcomes of recipients in the preferred model specification. DACA recipients increased the likelihood of working by 11.3 percentage points (p.p.) or 17.1% relative to non-participating non-citizens. With a total of 824,000 recipients, DACA moved 101,000 to 103,000 undocumented immigrants into employment in the 6 years following its introduction in 2012. This is driven by both movements into the labor force and out of unemployment. Pope (2016) estimated DACA moved 50,000 to 75,000 undocumented immigrants into employment in the first two years of the program. The effect of DACA on self-employment is an economically significant 18% decrease. As self-employment is used as a proxy for participating in the informal labor market, this shows DACA led to considerable shifts from the informal to the formal labor market. Receiving deferred action and work authorization also leads to an increase in school attendance by 3.4 p.p. or 13.5% increase from pre-DACA levels. Controlling for spillover effects in the alternative model, the effects of DACA participation are slightly larger with the effects of DACA on self-employment is now significant.

The improvement of labor market outcomes led to an increase in total income of 102.6% after receiving deferred action and work authorization through DACA relative to non-participating non-citizens. This increase is driven almost entirely by increases in wage and salary income which saw a 108.1% increase. Given the average pre-DACA total income of $15,117, this is an average increase in total income of $15,510. With an estimated 824,000 total DACA recipients, this amounts to a $12.8 billion increase in total income for the entire DACA participating population. In the alternative model, the effects of DACA on wages is a significant 7.5% increase relative to ineligible non-citizens.

The difference in magnitude between the the treatment-on-the-treated and intent-to-treat effects vary widely for each outcome of interest from 1.8 times higher for unemployment to 7.9 times higher for total income. These estimates vary considerably compared to the expected ratio of 2.61 under the assumption of no self-selection. Comparing the treatment-on-the-treated estimates from the preferred model to the intent-to-treat estimates using only the observed DACA-eligible indicator provides suggestive evidence that there may be considerable self-selection into the program. While the differences in the expected ratio may be due to self-selection, I cannot exclude that the differences are driven by heterogeneous effects across country-of-origin. I find that recipients from Asia had the largest labor market benefit from DACA. Latin Americans saw significantly lower labor market benefits compared to Mexican recipients but had the largest increase in school attendance from receiving deferred action and work authorization.

In the alternative model, eligible non-participants saw a 1.4 p.p. decrease in the likelihood of participating in the labor force relative to DACA ineligible non-citizens driven by unemployed eligible non-participants leaving the labor force. Eligible non-participants also saw a 2.3 p.p. decrease in school participation. The spillover effect of DACA on the total income for eligible non-participants was a statistically significant 30.3% decrease. This is driven by a 20.3% decrease in wage income, a 14.7% decrease in income from other sources, as well as an 11.7% decrease in hourly wage rate. Two important notes need to be made with regards to these estimates. First, as I can only estimate observed-eligible non-citizens this proxy is severely contaminated with authorized non-citizens which will attenuate the estimates towards zero. Second, these effects may be capturing eligible individuals attempting to become eligible but have not yet been approved at time t. The results may be driven by change in composition on who is classified as eligible non-participants over time. I control for this in an event study model by focusing on those eligible but never participated in DACA during the sample period.

A key assumption is that DACA recipients and eligible non-participants would follow the same trends as ineligible non-citizens if not for DACA. I construct a measure of ever participating in DACA to estimate an event study specification. I find insignificant pre-trends in support of the parallel trend assumption for most of the variables. When pre-trends are present, the trend is in the opposite direction of the effects of DACA. The event study specification also shows a decline in the effects of DACA on recipients after 2015.

To strengthen the validity of the assumptions made in the empirical models, I perform a number of robustness checks. First, The USCIS data also contains a total count of DACA recipients across state-of-residence and time. I construct an additional measure of DACA participation from the administrative data using variation in the number of DACA participants across state-of-residence. The results are robust to using this alternative proxy except for the effects on school attendance and hourly wage. I also perform a placebo test on naturalized citizens. If my constructed measures are only capturing DACA participation, then I should find no effect of DACA on this sample. The placebo test results are consistent with this assumption. A further robustness check takes advantage of a multiple proxy method proposed by Lubotsky and Wittenberg (Lubotsky and Wittenberg, 2006). This method includes both proxies simultaneously to minimize attenuation bias caused by proxy variables being mismeasured. The results are also robust to this alternative model.

This paper makes a number of important contributions to the emerging literature on the impact of DACA. First, I use administrative data to create a probability measure of DACA participation to estimate the treatment-on-the-treated effects of DACA. I also estimate the spillover effects DACA has on eligible non-participants. Second, I estimate the effect of

DACA across different sources of income rather than total income alone. Desegregating the effect DACA has on total income across different income sources provides a better understanding on how DACA is affecting the labor market outcomes of recipients. It also of value to public policy if the increase in total income is driven by taxable income (wages and salaries), likely informal income (income from other sources), or if amnesty burdens the welfare system (income from welfare). Lastly, I contribute to the literature on DACA by expanding the number of years included in the sample to include the Trump presidency which led to considerable changes and legal challenges to the DACA program.

This work also relates to the work estimating the effects on the 1986 Immigration Reform and Control Act (IRCA). IRCA granted 2.8 million undocumented immigrants amnesty and a pathway to citizenship (Baker, 2015). Most studies have found IRCA lead to an increase in participants income (Kossoudji and Cobb-Clark, 2000; Amuedo-Dorantes et al., 2007; Lozano and Sorensen, 2011). However, Amuedo-Dorantes et al. (2007) found a decrease in labor force participation and an increase in the unemployment.

This paper also relates to the strand of literature attempting to identify unobservable populations in survey data using secondary sources. Lozano and Sorensen (2011) predict undocumented status in the Census data using the Mexican Migration Project Survey which does ask about Mexican immigrants documentation status. Other work has focused on using the Survey of Income and Program Participation to predict undocumented status into Census surveys (Van Hook et al., 2015). Bollinger and Hagstrom (2008) and Bollinger and Hagstrom (2011) use administrative data to predict refugee status in the Current Population Survey. I expand on this work by using administrative data to create a probability measure of DACA participation in the ACS. As asking about legal status becomes a more sensitive status, identifying undocumented immigrants will become more difficult over time. Using administrative data will become an important method to estimate the effects of policy on these unobserved populations.

The paper continues as follows. Section 2.2 details the institutional framework of DACA. Section 2.3 describes the empirical strategy followed by Section 2.4 which discusses the data and variable construction. Results of the effect of DACA on the labor market outcomes of DACA recipients are presented in Section 2.5. Section 2.6 provides estimates of the spillover effect on eligible non-participants. Estimates from an event study framework are presented in Section 2.7. Section 2.8 presents results of the robustness checks using an alternative measure and model. Finally, Section 2.9 provides concluding remarks.

2.2 Deferred Action for Childhood Arrivals

On June 15, 2012, through prosecutorial discretion, President Obama enacted an executive memorandum announcing the Deferred Action for Childhood Arrivals program. This executive decision was taken after the failed attempts in Congress to pass the DREAM Act in 2010 and 2011 which would have provided permanent residency to undocumented immigrants that came to the US as children. The DACA program is arguably the largest immigration reform impacting undocumented immigrants since the passage of the Immigration Reform and Control Act (IRCA) in 1986 where nearly 2.7 million undocumented immigrants were approved for permanent residency (Rytina, 2002). The memorandum provides eligible applicants with two key benefits. First, recipients receive two years of relief from deportation. Second, recipients receive work authorization through an Employment Authorization Document (EAD). With the EAD, individuals are allowed to apply for a Social Security Number (SSN). Both benefits are subject to renewal every two years. DACA does not provide any form of legal immigrant status or a pathway to citizenship. This is a *de facto* temporary legalization for the participating population.

On September 5, 2017, President Trump ordered the termination of the DACA program. Court challenges have led to a continuation of renewals for those who have already been approved prior to the termination of DACA but no new applications have been accepted since September 5, 2017. Following these legal challenges, the Supreme Court ruled that the administration acted arbitrarily and capriciously in ending the program and ruled that the rescission be vacated. After the 2020 presidential election, President Biden reinstated DACA in full.

2.2.1 Requirements for DACA Eligibility

The DACA program is specifically targeted towards undocumented immigrants that came to the US as children. In order to qualify for DACA, an applicant must meet the following six criteria:

1. have no lawful status as of June 15, 2012;
2. have come to the United States before the age of 16;
3. have been under the age of 31 as of June 15, 2012;
4. have continuously resided in the United States since June 15, 2007 and be physically present on June 15, 2012;
5. must be currently in school, have completed high school (or have obtained a General Education Development (GED)) certificate, or be an honorably discharged veteran of the United States military;[1]
6. cannot have been convicted of a felony, significant misdemeanor, or three or more other misdemeanors.

In addition to these requirements, an applicant has to pay a processing fee of $465 and be at least 15 years of age.

2.2.2 Size of the Eligible Population

In 2012, the Pew Research Center estimated 1.7 million of the 11.2 million undocumented immigrants are potentially eligible for DACA (Passel and Lopez, 2012). Of these, 950,000 were immediately eligible (satisfy criteria 1, 2, 3, 4, 5 and age 15 and older) at the announcement of DACA. Another 450,000 were potentially eligible by aging into the program (between the ages of 5 to 14). Given the age distribution of the those 5 to 14 years old, the number of children aging into the eligibility annually range from 80,000 to 90,000 between

[1] Very few DACA-eligible immigrants have used military service to satisfy this requirement. Approximately 820 DACA recipients are in the military (USCIS, 2017a)

2013 and 2016 (Batalova et al., 2014). Another 320,000 are potentially eligible by completing the education requirements (those aged 16 to 30 who do not have a high school diploma or GED and are not enrolled in school).

The above estimates of the number of potentially DACA-eligible immigrants does not take into account criteria 6 (not having been convicted of a felony or three serious misdemeanour) as neither the CPS nor ACS asks about the criminal history of participants. The Migration Policy Institute (MPI) estimates that 820,000 of the 11 million undocumented immigrants had criminal convictions (Rosenblum, 2015). Of those 820,000, Rosenblum (2015) estimate approximately 300,000 had a felony conviction (a conviction with a sentence 1 year of longer) and 390,000 had serious misdemeanours (a conviction with a sentence 90 days or longer) making them ineligible for DACA even if they satisfied all the other requirements. No information is available on the number of undocumented immigrants that have three or more misdemeanours, though, Rosenblum (2015) estimates at most 130,000 have at least one misdemeanour.

2.2.3 Size and Characteristics of DACA Recipients

The USCIS began accepting applications for DACA on August 15, 2012. Figure 2.1 displays the cumulative number of initial approvals from implementation in 2012 to the end of 2018 by quarter. Figure 2.2 shows cummulative number of initial approvals for Mexicans (a) and top 25 country-of-origin groups excluding Mexico (b). Data comes from publicly available quarterly reports published by the USCIS from 2013 to 2018 (USCIS, 2017b, 2018). By the end of 2013, the first full year of the program, 520,000 applications had been approved. The rate of new applications slowed beginning in 2014 and as of 2018 an estimated 824,000 undocumented immigrants have received deferred action and work authorization through DACA. A total of 80,400 applicants have had their initial applications denied. This is a take up rate of roughly 50% out of the estimated 1.7 million potentially eligible. Receiving work authorization through an EAD universally led to receiving a SSN. From October 1st, 2012 to June 30th, 2017, the Social Security Administration assigned original SSNs to 838,058 non-citizens that had been granted DACA status (SSA, 2018).[2]

A total of 14,434 renewal applications have been denied by the end of 2018 (USCIS, 2017b). Investigations from the House and Senate Judiciary Committees in 2017 estimate 40,000 DACA participants were able to have their legal status changed as of August 2017 (Grassley, 2017). The Congressional investigations estimated 59,778 DACA recipients have applied for Lawful Permanent Resident (LPR) status through Advance Parole and 39,514 have been approved (4.8% of total DACA recipient) (Grassley, 2017). Of those who received LPR status, 2,181 have applied for U.S. citizenship and 1,056 have become U.S. citizens (Grassley, 2017). While DACA does not provide a formal "pathway to citizenship," DACA participants can take advantage of Advance Parole through an I-131 application for travel documents. Advance Parole is a formal procedure that allows DACA participants the ability to leave the country and then formally request an immigration status change in the US embassy in their home country. Without Advance Parole, an undocumented immigrant who leaves the U.S. will be banned from reentering the country for 3 or 10 years depending on how long they have been in the country without authorization. Given that 824,000 applicants were approved, 14,000 have had their status terminated, and 40,000 have transitioned to LPR, an estimated 92,000 (11%) DACA recipients have let their DACA status expire by the end of 2018. As of the end of 2018, there were approximately 680,000 active DACA recipients (USCIS, 2019).

The composition and geographic dispersion of DACA applicants is similar to that of the undocumented immigrant population as a whole. As of 2018, around 92% of approved applicants are from Central and South America and 76% are from Mexico alone (USCIS, 2017b). California and Texas account for over 229,000 and 127,000 of the initial approved applicants, respectively. Illinois, New York, and Florida have around 40,000 each (USCIS, 2017b, 2018). These five states make up 52% of the total number of applicants. Of those with active status as of November 2018, 47.3% are male, 79% are single, and are, on average, 24 years old (USCIS, 2019).

[2]The USCIS states that there have been a total of 792,000 application approvals as of June 30th, 2017 (USCIS, 2017b) leading to a discrepancy of 46,000 applicants between the number of DACA approvals given by the USCIS and the number of SSNs given to DACA recipients by the SSA.

2.3 Model Specification and Empirical Strategy

In an ideal situation, the ACS would have an indicator that would identify whether and when an individual was a recipient of deferred action and work authorization through DACA. Using a sample of non-citizens ages 18-35 with at least a high school degree, I would estimate the following reduced-form model specification:

$$Y_{icst} = \beta_0 + \beta^* Recipient_{icst} + \beta_2 X_{it} + \beta_3 W_{it} + \beta_4 Z_{st} + \gamma_c + \gamma_t + \gamma_s + \gamma_s t + \epsilon_{icst} \tag{2.1}$$

where Y_{icst} is the outcome of interest for non-citizen i from country c in state s in year t. The indicator $Recipient_{icst}$ takes the value of one if non-citizen i was ever a DACA recipient by time t and zero otherwise. The vectors X_{it}, W_{it}, and Z_{st} contain key demographic and state-by-year controls. The vectors γ_c, γ_t, and γ_s allow for country-of-origin fixed effects, time fixed effects, and state fixed effects, respectively. Lastly, $\gamma_s t$ allows for state-specific time trends.

The coefficient of interest, β^*, represents the treatment-on-the-treated effect of DACA. The control group would be all non-citizens in the sample that did not participate in the DACA program. This group contains both DACA ineligible non-citizens and Eligible non-participants. The coefficient, β^*, therefore represents the change in the outcome of interest after an individual becomes a DACA recipient relative to all non-citizens who did not participate in the DACA program. The key assumption to get unbiased estimates of β^* is that both the treated and control group would have followed similar trends had it not been for the enactment of DACA.

Unfortunately, large nationally-representative datasets commonly used by researchers, such as the ACS, do not ask whether an individual participated in DACA. As $Recipient_{icst}$ is unobserved, I cannot directly estimate the effect of DACA on recipients (β^*). To deal with this limitation, I consider the conditional expectations model of equation 2.1:

$$E[Y|C,T,E,X,W,Z,S] = \beta_0 + \beta^* E[R|C,T,E,X,W,Z,S] + \beta_2 X + \beta_3 W + \beta_4 Z + \gamma_c + \gamma_t + \gamma_s + \gamma_s t \tag{2.2}$$

Where R represents DACA participation. The expected conditional value of R can be represented as:

$$E[R|C,T,E,X,W,Z,S] = P(R=1|C,T,E,X,W,Z,S) \tag{2.3}$$

Using publicly available data from the DHS on the total number of DACA recipients by country of birth in each year along with estimates of the total DACA eligible population using the ACS, I am able to construct the probability a non-citizen was ever a DACA recipient at time t.

$$P(R=\widehat{1|C},T,E) \tag{2.4}$$

The variable $P(R=\widehat{1|C},T,E)$ represents the probability a non-citizen i from country c had ever participated in DACA by time t. The probability is set to zero if the non-citizen does not meet any one of the observable DACA-eligibility requirements (E = 0) or if the year is 2012 or earlier.[3] The procedure of constructing this probability is detailed in Section 2.4. Under the assumption that the constructed probability is equal to the expected mean of the unobserved DACA participation or,

$$P(R=\widehat{1|C},T,E) = P(R=1|C,T,E,X,W,Z,S) = E[R|C,T,E,X,W,Z,S] \tag{2.5}$$

I can plug in variable 2.4 into equation 2.2 and get the preferred model specification to be estimated in this paper:

$$E[Y|C,T,E,X,W,Z,S] = \beta_0 + \beta^* P(R=\widehat{1|C},T,E) + \beta_2 X + \beta_3 W + \beta_4 Z + \gamma_c + \gamma_t + \gamma_s + \gamma_s t \tag{2.6}$$

The assumption in equation 2.5 underlying the preferred model presumes the constructed probability is only capturing the unobserved participation decision. Specifically, that there are no unobserved time varying country-of-origin factors that are correlated with the unobserved participation decision and with outcomes. With this assumption the probability measure constructed is an instrument for the unobserved treatment ($recipient_{icst}$). In a linear regression model, as the probabilities are instruments for the true unobserved treatment, the probability measure can be used as a regressor (Bollinger and Hagstrom,

[3]While the program was announced in June 2012 and applications submitted by the end of 2012, very few applicants were approved in 2012. Additionally, the publicly available data does not provide the country-of-origin of those that were approved in 2012.

2011). Under the assumption in equation 2.5 and the assumption that both treated and control groups would have similar trends if not for DACA equation 2.6 provides an unbiased estimate of β^*.

The sample in the preferred model of equation 2.6 comprises of all non-citizens ages 18 to 35 with at least a high school degree (or equivalent). The outcomes of interests are labor force participation, the likelihood of working, of being unemployed, the usual hours worked in a week, and likelihood of being self-employed. I look at total income and its subcomponents. The specific subcomponents are income from wage and salaries, welfare income, and income from other sources[4]. I also look at wage of employed individuals with positive hours of work. Additionally, I look at the likelihood of currently attending school.

The vector X contains demographic controls including education, sex, race, ethnicity, and marital status. The vector W includes fixed effects for individual $i's$ age and age when they arrived in the United States to control for the eligibility criteria. Vector Z is a vector of state-by-year controls. This vector includes the state unemployment rate, an indicator if the state has enacted universal E-Verify, if the state has enacted public E-Verify, an indicator if the state grants in-state tuition to undocumented immigrants, and an indicator if the state allows undocumented immigrants to obtain a driver's license at time t. The vector γ_c allow for country-of-origin fixed effects. The vectors γ_t and γ_s allow for time and state fixed effects, respectively. Lastly, $\gamma_s t$ allows for state-specific time trends. When estimating Equation 2.6, standard errors are clustered at the state-by-year level.[5]

A one unit change in the variable $P(\widehat{R = 1|C}, T, E)$ corresponds to a shift in the probability of ever receiving DACA in year t from zero to one. The coefficient β^* therefore corresponds to the effect of having had received DACA by year t on the outcomes of interest relative to non-participating non-citizens. Individuals that had received DACA may no longer have active status at time t and would no longer have the accompanying benefits. Due to this, the estimates of β^* will be a lower bound of the effects of DACA on active DACA recipients.

[4]Income from other sources are income not from wages, welfare, farm or business/investments.

[5]Results are robust to clustering standard errors at the state level, and at the country-of-origin level.

2.4 Data

To analyse the labor market effects of DACA participation, I use the ACS sourced from IPUMS (Ruggles et al., 2020). The main sample used is comprised of all non-citizens aged 18 to 35 with at least a high school degree (or equivalent) between the years 2005 to 2018. The choice of years was dictated by survey availability. The first year for which the nationally representative data were collected was 2005 and 2018 was the most recent sample available. The ACS contains key labor market outcomes including labor force participation, employment, unemployed, their usual hours worked, and their self-employment status. Information on individual's total income and by source. Sources analyzed in this paper are income from wages and salary, income from welfare, and income from other sources.

The ACS also contains important demographic and immigration related variables which are used to determine respondents' observable DACA eligibility status and assign a probability of DACA participation such as each immigrant's country-of-origin, their U.S. citizenship status, number of years spent in the U.S., quarter of birth, and educational attainment. The ACS does not contain data on non-citizens legal status and therefore undocumented status is proxied using non-citizenship. As is done in Pope (2016) and Giuntella and Lonsky (2020), I define non-citizens as being DACA-eligible as those who: (1) were under the age of 31 as of June 15, 2012; (2) have lived in the U.S. since June 15, 2007; (3) entered U.S. before reaching 16th birthday; (4) have at least a high school degree (or equivalent); (5) were born outside the U.S. or its territories; (6) are not U.S. citizens; and (7) at least 15 years of age.[6]

The key variable is the probability a non-citizen is a DACA recipient at time t. To construct this variable, I take advantage of administrative data from the USCIS that provides a total count of DACA recipients by country-of-origin and year (USCIS, 2017b, 2018). The USCIS data provides a total count for all countries only for the year 2018. For the years before 2018, only the 25 largest country-of-origin groups are observed. Using the 2018 total count, I do a linear extrapolation to estimate the total number of DACA recipients from

[6] To define the DACA-eligible population in year 2012 and before, the criteria are restricted to non-citizens who were: (1) under the age of 31 as of June 15 of the previous calendar year; (2) arrived to the U.S. prior to their 16th birthday, (3) have lived in the U.S. for at least 6 years, (4) have at least a high school degree (or equivalent); (5) were born outside the U.S.; and (6) are not U.S. citizens.

countries not observed in the years 2013 to 2017 beginning at zero in 2012.[7] I use the ACS to estimate the average count of DACA-eligible non-citizens by country-of-origin from 2013 to 2018. Using the count provided in the administrative data and the estimate of the size of the observed DACA-eligible population, I construct an estimate of the probability a non-citizen i from country c at time t is a DACA recipient as follows:

$$P(\widehat{R = 1|C}, T, E) = \frac{(\text{Total DACA Recipients})_{ct}}{(\text{Total DACA Eligible Population})_{c}} \cdot Eligible_{it} \tag{2.7}$$

where the numerator is constructed from the USCIS data and the denominator is constructed from the ACS data. The average number of DACA-eligible non-citizens for each country-of-origin across all post DACA years is used to limit sampling error which can be an issue with nationalities that have small populations. Total DACA recipients is equal to zero in the pre-DACA years 2005 to 2012.[8] The indicator $Eligible_{it}$ is equal to one if an immigrant in the sample meets all observable DACA requirements and zero otherwise. For individuals who are not observed to be DACA-eligible, the probability of being a DACA recipient is set to zero. Each individual in the sample is assigned a probability of being a DACA recipient based on their observed eligibility status, country-of-origin, and year observed in the sample.

The administrative data also provides a total count of DACA recipients by state-of-residence for each year since the enactment of DACA up to the year 2018. Data is available for all states in all years. The same procedure as above is done to construct the probability measure of DACA participation across state-of-residence (S). I construct an estimate of the probability an observed DACA-eligible non-citizen i from state-of-residence s at time t is a DACA recipient as follows:

$$P(\widehat{R = 1|S}, T, E) = \frac{(\text{Total DACA Recipients})_{st}}{(\text{Total DACA Eligible Population})_{s}} \cdot Eligible_{it} \tag{2.8}$$

Each individual in the sample is assigned a probability of being a DACA recipient based on their observed eligibility status, state-of-residence, and year observed in the sample.

[7] Results are robust to using a sample of only the top 25 largest country-of-origin groups.

[8] While there was in fact a few number of approvals in the year 2012, they were at the very end of the calendar year and the total count publicly available are not desegregated at country-of-origin level.

This creates two proxies for DACA participation using different sources of variation. The proxy $\widehat{P(R = 1|C, T, E)}$ is the probability a non-citizen had ever participated in DACA by time t using variation across country-of-origin while $\widehat{P(R = 1|S, T, E)}$ is the probability a non-citizen had ever participated in DACA by time t using variation across state-of-residence.

2.4.1 Summary Statistics

Table 2.1 provides descriptive statistics of the constructed probability an observed DACA-eligible non-citizen ever participated in DACA for the 20 largest DACA approved countries in the year 2018. At a total of 643,373 approvals, Mexicans make up the overwhelming share of DACA recipients with 78% of all approvals. Observed DACA-eligible Mexicans have the largest share of DACA recipients at 71%. There is significant variation across countries and within regions. The share of DACA recipients in Asian countries range from 10.9% for Filipinos, Indians (11.8%), Chinese (3.4%), Koreans at 21.5%, and 24.1% for Indonesians. With 13% share of DACA-eligible non-citizens participating in DACA, Poland is the only European country in the top 30 list. For African countries, 20% of observed DACA-eligible Nigerians and 17.3% of Kenyans participated in DACA. With 990 approvals, only 3.2% of DACA-eligible non-citizens from Canada participated in DACA. There is a zero probability of an observed DACA-eligible non-citizen that migrated from England or Scotland is a DACA participant.[9] Figure 2.3 shows variation in the probability an observed DACA-eligible immigrant is a DACA recipient across time for the largest 25 country-of-origin groups.

The main source of variation in the measure comes from differences in the share of undocumented immigrants within each country's observable DACA-eligible population. Countries that have a high share of legal immigrants, such as Western European countries, China, Cuba, and Canada will have lower probability measures. Countries with a high share of undocumented immigrants, such as Mexico, Central and South American countries will have a higher probability measure. Additional variation is caused by differences in the

[9] Descriptive statistics of the constructed probability an observed DACA eligible non-citizen ever participated in DACA using variation across state-of-residence from equation 2.8 can be found in Table A.1.

participation rates across countries.

Table 2.2 provides descriptive statistics of the DACA recipients, eligible non-participants, ineligible non-citizens, and the full sample during the pre-DACA period (years 2005 to 2012). To produce summary statistics of the demographic characteristics of the DACA recipient population (column 1) prior to the enactment of DACA I assign the constructed probability a non-citizen ever participated in DACA by the year 2018 to all pre-DACA years. This probability is then multiplied by the ACS provided person-weight. To get descriptive statistics of the eligible non-participating population (column 2) I multiple the person weight by the probability a DACA eligible non-citizen was never a DACA participant.[10] To get descriptive statistics of the ineligible non-citizen population (column 3) I restrict the weighted sample to those that have the eligiblity indicator equal to zero. Column 4 shows the weighted summary statistics of the full sample. Recipients compose of 8.6% of the weighted sample. This equates to DACA recipients making up 57.2% of the observed DACA-eligible sample population. This values are similar to estimates of the share of DACA participants for the respective populations provided by Hipsman et al. (2016) and Passel and Lopez (2012).

Table 2.2 shows that, relative to eligible non-participants, DACA recipients are less likely to be in school (25.1% and 37.16%) and less likely to have a college degree (5.58% and 13.09%). Almost 92% of recipients are Hispanic, driven by 78% of recipients being from Mexico. Observed DACA-eligible non-participants are more likely to be from Latin America (31.51%), Asia (23.86%) and Europe (12.42%). The demographics are consistent with the observed DACA non-participating population having a significant proportion of documented immigrants. DACA-ineligible non-citizens are more likely to be recent immigrants with only 6.35 years living in the US compared to 15.7 years in the US for DACA recipients. Ineligible non-citizens are more likely to have entered the country as adults (22 years old vs. 8.5 years old) and are twice as likely to be married than eligible non-citizens.

In Table 2.2, a number of differences are shown in each group's labor market outcomes pre-DACA. While DACA recipients are 2.6 percentage points more likely to work and work 1.3 more hours a week than their non-participating eligible counterparts, they earn

[10] This is calculated by the equation non-participant = (1 - $P(recipient)_{ic,2018}$)* Eligible$_{it}$.

roughly \$1,300 less in total income. DACA-ineligible non-citizens have considerably higher total income at \$23,929 or 1.6 times larger than DACA-participants. DACA-ineligible non-citizens are also 3.7 percentage points less likely to be unemployed, work an hour more a week, and 1.5 percentage points more likely to be self-employed.

2.5 Results

The estimates of the preferred model from equation 2.6 of the effect of DACA on labor market outcomes of DACA recipients are reported in Table 2.3. Row 2 of Table 2.3 show the pre-DACA means of the labor market outcomes for DACA recipients. The results reveal DACA has significantly improved the labor market outcomes of recipients relative to the control group. In the preferred model, the control group is composed of both DACA-eligible non-participants and DACA-ineligible non-citizens.

Column 2 of Table 2.3 shows that DACA recipients are 11.3 p.p. more likely to be working compared to the control group after receiving deferred action and work authorization through DACA. With an estimated 66% of DACA recipients being employed prior to 2012, the estimates translate to an increase in the likelihood of working of 17.1%. This is a little less than 3 times larger than the attenuated intent-to-treat effects of DACA on the observed DACA eligible population in the first two years of the program estimated in Pope (2016). The increase in the likelihood of working is driven by individuals entering the labor force and individuals moving out of unemployment. Column 1 shows DACA increased the likelihood a DACA recipient is in the labor force by 9.6 p.p. while Column 3 shows that DACA recipients decreased the likelihood of being unemployed by 3.0 p.p.. Column 4 indicates DACA recipients work 3.802 more hours a week relative to the control group. This outcome is an alternative measure for working. The estimates can be viewed as DACA leading to one additional full-time job (40 hours per week) per 10.5 DACA recipients. For reference, Pope (2016) estimated an increase of one additional full-time job per 23 DACA-eligible individuals. These estimates indicate lack of work authorization and fear of deportation are severe frictions limiting undocumented immigrants from entering the labor force and acquiring employment when they do enter. The effects of DACA on self-employment is statistically significant and the magnitude of the coefficient indicates a 18% decrease in the likelihood of self-employment. As self-employment is more likely to indicate employment in the informal sector this indicates economically and statistically significant movement from the informal to the formal labor market.

I also estimate the effect of DACA on recipient's school attendance. Column 6 of Table 2.3 shows DACA recipients are 3.4 p.p. (13.5%) more likely to be enrolled in school relative to the control group. This is in contrast with the difference-in-differences results from Pope (2016), Amuedo-Dorantes and Antman (2017), and Hsin and Ortega (2018) on the effect of DACA on DACA-eligible non-citizens but consistent with the results from Kuka et al. (2020) and Ballis et al. (2020).

Estimates from the preferred model of the effect of DACA on the income of DACA recipients are reported in Table 2.4. Row 2 of Table 2.4 shows the pre-DACA means of income for DACA recipients. The income outcomes are transformed by their inverse hyperbolic sine so that the coefficients are interpreted as percent changes. A benefit of this transformation rather than a simple log-transformation is that I am able to include those in the sample with zero income.

The results reveal DACA has significantly increased the total income of DACA recipients relative to eligible non-participants and ineligible non-citizens. Column 1 in Table 2.4 shows that DACA recipients saw an increase in total income of 102.6% after DACA relative to the control group. The pre-DACA income of DACA recipients was $15,117, indicating DACA lead to an increase in total income of $15,510 for the average DACA recipient. Pope (2016) did not find a significant average effect of DACA-eligibility on total income but did note an increase of 5-20% in total income for those the income distribution between the 30th and 60th percentile, or around an increase of 400 to 800 dollars.

The increase in total income is driven almost entirely by a 108.1% increase in income from wage and salaries (Column 2). DACA had no effect on income from other sources (Column 3). Column 5 of Table 2.4 shows the effect of DACA on recipients hourly wage. Hourly wage is constructed by diving wage income by usual hours worked in a week times weeks worked in a year. The inverse hyperbolic sine transformation is also taken so that the estimated effects translate to percent changes. I find DACA recipients had a statistically insignificant 1.6% increase in wages relative to the control group.

While undocumented immigrants, including DACA recipients, are ineligible to participate in federal welfare programs, some states do provide undocumented immigrants access to certain state funded welfare programs. Column 4 shows the estimates of DACA on recipient's welfare income. DACA did not have a statistically significant effect on welfare income of DACA recipients. Extrapolating this estimate to future amnesty programs should be taken with caution though as participating in a permanent amnesty program will provide participants access to federal welfare programs and as such may lead to an increase in welfare expenditure.

2.5.1 Relationship To Intent-to-Treat Estimates

To be able to extrapolate the average treatment-on-the-treated effects from the intent-to-treat estimates in the literature we must make a number of strong assumptions. First, due to data limitations, the literature on DACA has focused on estimating the effects of being an observed DACA-eligible non-citizen as undocumented status (and criminal history) is unobserved. The observed DACA eligibility indicator is contaminated with authorized (DACA-ineligible) non-citizens. This will cause the estimates of the intent-to-treat effects to be attenuated towards zero. Assuming DACA does not have an effect on observed eligible non participants, the degree of attenuation can be characterized by the following equation:

$$\widehat{\beta^{ITT}} = \beta^{ITT} \cdot P(E^* = 1|E = 1) + 0 \cdot P(E^*1 = 0|E = 1) \tag{2.9}$$

or,

$$\beta^{ITT} = \widehat{\beta^{ITT}} \cdot \frac{1}{P(E^* = 1|E = 1)} \tag{2.10}$$

Where E represents observed eligible and E^* represents the true unobserved eligibility status. The coefficient β^{ITT} is the true intent-to-treat effects and the coefficient $\widehat{\beta^{ITT}}$ is the attenuated intent-to-treat estimate from the observed eligibility indicator. The scaling factor to get the true intent-to-treat from the attenuated estimate is 1 over the probability an observed eligible noncitizen is actually eligible. Disregarding criminal history, the scaling factor will be 1 over the probability an observed DACA eligible non-citizen is actually an undocumented immigrant.

Next, to extrapolate the treatment-on-the-treated effects from the intent-to-treat effects we must assume that (1) there is no self-selection into the participation decision and (2) there are no spillover effects. Note that the intent-to-treat effects can also be written as;

$$\beta^{ITT} = \beta^* \cdot P(R = 1|E^* = 1) + 0 \cdot P(R = 0|E^* = 1) \tag{2.11}$$

Where R represents DACA participation. The probability$P(R = 1|E^* = 1)$ is the probability a DACA-eligible immigrant participated in DACA. The probability$P(R = 0|E^* = 1)$ is the probability a DACA-eligible immigrant did not participate in DACA. The coefficient, β^*, is again the treatment-on-the-treated effects of DACA. Plugging in equation 2.10 into equation 2.11, we get the relationship between the attenuated intent-to-treat estimates and the assumed treatment-on-the-treated effects if the above assumptions hold.

$$\widehat{\beta^{ITT}} \cdot \left[\frac{1}{P(E^* = 1|E = 1)} \cdot \frac{1}{P(R = 1|E^* = 1)} \right] = \beta^* \tag{2.12}$$

Equation 2.12 shows that the attenuated intent-to-treat estimates must be scaled by 1 over the probability an observed DACA eligible non-citizen is actually an undocumented immigrant times 1 over the participation rate of DACA eligible immigrants. Estimates from the DHS indicate the share of unauthorized immigrants among the non-citizen population aged 18-35 is around 61% (Baker and Rytina, 2013), indicating the attenuated estimates need to be scaled by 1.64 to derive the true intent-to-treat estimates. As an estimated 63% of DACA-eligible individuals participated in the DACA program (Hipsman et al., 2016), the assumed treatment-on-the-treated effects should be 1.59 times larger than the true intent-to-treat effects. In other words, under the assumption of no self-selection into the program and no spillover effects, the treatment-on-the-treated effects should be (1.64*1.59 =) 2.61 times larger than the attenuated estimates using the observed eligibility indicator.

I now compare the treatment-on-the-treated estimates in this paper to those of the attenuated intent-to-treat estimates using the observed eligibility indicator. A benefit of this exercise is that it will provide some evidence of whether self-selection is taking place among DACA recipients. If the assumptions of no self-selection and no spillover effects hold, the ratio between the two estimates should be around 2.61. Although it provides suggestive evidence, I can not rule out other possible mechanisms, such as heterogeneous treatment effects across country-of-origin as a possible source if the scaling factor is not equal to 2.61.

To do the comparison, I estimate a similar difference-in-differences model as in Pope (2016) using the same sample and controls as in equation 2.6.[11]

$$\begin{aligned} Y_{icst} = \beta_0 + \beta_1 eligible_{it} \cdot post_t + \beta_2 eligible_{it} \\ + \beta_3 X_{it} + \beta_4 W_{it} + \beta_5 Z_{st} + \gamma_t + \gamma_s + \gamma_s t + \epsilon_{icst} \end{aligned} \tag{2.13}$$

The coefficient , β_1 , is the effect of DACA on DACA-eligible non-citizens relative to ineligible non-citizens. This captures the attenuated intent-to-treat effects of DACA ($\widehat{\beta^{ITT}}$). All other controls are the same as in equation 2.6 so as to make sure the results are not driven by differences in model specification or sample selection. Standard errors are clustered at the state-year level.

Table 2.5 reports the effect of observed DACA eligibility on labor market outcomes. The second row shows the ration between the estimated treatment-on-the-treated effects and the estimated attenuated intent-to-treat effects. These estimates compare closest to those in Table 3 Row C in Pope (2016). The Magnitude of the attenuated intent-to-treat effects are lower compared to Pope's estimates. This is driven by my sample having six post DACA years compared to Pope (2016) which only estimated the effect of DACA on the first two years of the program. The model specification also includes country-of-origin fixed effects and state-year immigration controls not included in Pope (2016).

The treatment-on-the-treated effects are considerably larger and vary widely relative to the attenuated intent-to-treat effects. The effects of DACA on recipient's labor market outcomes range from 1.88 larger than the attenuated intent-to-treat (likelihood of being unemployed) to 5.66 times larger (usual hours worked). The treatment effects of DACA on the likelihood of working are 3.77 times larger and 5.05 times larger for likelihood of being in the labor force than the intent-to-treat effects. Looking at the effect of DACA eligiblity

[11] Of course, this is not an exact comparison. In the preferred model, the estimated treatment-on-the-treated effects are the contemporaneous effects of participating in DACA at time t. In the difference-in-differences model, the intent-to-treat effects are the average effects in all post DACA years.

on schooling would give insignificant results. The effect on recipients is 4.25 times larger and significant. The scaling factor for self-employment and schooling are 3.00.

Table 2.6 reports the effect of observed DACA-eligibility on income. The second row shows the estimated treatment-on-the-treated effects over the estimated attenuated intent-to-treat effects. The treatment effects of DACA on recipient's total income is 7.87 times larger than when using the observed eligiblity indicator. The scaling factor for the effects on wage and salary income 5.60, while it is 0.13 for income for other sources, and 2.30 for welfare income. The effect of DACA-eligiblity on hourly wage is of opposite sign and significant at a 6% decrease giving a scaling factor of -0.26.

This results demonstrate that it is difficult to extrapolate the treatment-on-the-treated effects from intent-to-treat estimates and therefore needs to be empirically estimated directly as is done in this paper. A possible reason for this is that the assumption of no self-selection into the program may not be valid. Although, I also cannot rule out other possible mechanisms such as heterogeneous treatment effects.

The estimated intent-to-treat effects and the treatment-on-the-treated effects are both of value to policy markers as they answer separate, though interrelated, questions. In this paper, I have estimated the effects of ever participating in DACA directly. This provides valuable information to policy makers when estimating DACA's cost and benefits as they are driven by those that participate in the program. The intent-to-treat effects are of value as it provides the effects of policy parameters which policy makers can actually change when formulating policy. As shown though, the intent-to-treat estimates may not be informative in understanding the effects of amnesty programs on actual participants and why it is important to measure those effects directly.

2.5.2 Heterogeneous Effects Across Region-of-Birth

The difference between the treatment-on-the-treated Estimated from the preferred model and those of the intent-to-treat estimates using the observed eligbility indicator may be a result of heterogeneous effects of DACA across country of origin. Such that individuals from countries whose observed DACA-eligible non-citizen population have a higher probability of participating in DACA also benefit the most from participating in DACA. I next estimate equation 2.6 where the probabilistic measure is interacted with indicators for region-of-birth. The regions considered are Latin America (Excluding Mexico), Asia, Europe, and all other countries (Rest).

Table 2.7 and Table 2.8 shows the heterogeneous effect of DACA on recipients. As the excluded group is Mexicans, the coefficient on $P(R|C, T, E)$ measure the effect of DACA on Mexican Recipients. The coefficients on the interaction terms are the relative effects compared to Mexican recipients. Significant Heterogeneity in the effect of DACA is documented.

Recipients from Asian countries have the largest documented treatment-on-the-treated effects on labor market outcomes and Income. Latin Americans have significantly lower labor market outcomes compared to Mexicans but have the highest increase in school attendance. This result shows even within ethnicity, considerable heterogeneous effects are at play. The effects on Europeans and Rest of countries are mostly statistically insignificant or have unrealistic coefficient estimates. This is likely due to Europeans and Rest making up less than 1% of the DACA participating population each.

These results shows a significant limitation in the previous literature. Past work has focused on estimating the effects of DACA on a sub-sample of Hispanics only or Mexicans only as those groups have the highest share of DACA participants across their non-citizen populations. As non-Hispanics make up a small fraction of DACA participants, prior work has not focused on them. Focusing on only Hispanics misses important race-ethnicity heterogeneity that needs to be further studied. Asian DACA-participants appear to benefit the most from conditional amnesty provided by DACA. This is important for policy makes as Non-Hispanics continue to make up a greater share of the undocumented population (Passel and Cohn, 2019). Further work needs to examine why Asian undocumented immigrants have the highest benefits from DACA but yet have some of the lowest participation rates (Hipsman et al., 2016). These results also indicate that the variability in the ratio between the treatment-on-the-treated estimates and the intent-to-treat estimates are likely caused by heterogeneous effects. Whether and to how much self-selection plays a role needs further examination.

2.6 Spillover Effects on Eligible Non-participants

An estimated 37% of DACA eligible immigrants have not applied for deferred action and work authorization through DACA (Hipsman et al., 2016). The announcement of DACA and the regularization of 824,000 similar undocumented immigrants may have impacted this eligible non-participating group. This spillover can be caused by increased competition driven by the large increase in labor force participation documented in Section 2.5. The announcement of DACA may also have led to behavioral changes if it altered their expected probability they are staying in the country or if those individuals want to apply but cannot afford the financial and legal costs. Spillover effects on this population will alter the cost and benefits of amnesty and regularization programs such as DACA and need to be analyzed. Controlling for the effect of DACA on eligible non-participants will also more accurately estimate the effect of DACA participants. The control group in the preferred model of equation 2.6 is composed of two distinct populations, non-citizens eligible for DACA but did not participate and non-citizens that are ineligible for DACA. If there is in fact significant spillover effects on eligible non-citizens (control group) then the estimated results from Section 2.5 will be biased.

Given the policy importance of estimating spillover effects to understand the total effect of DACA and to more accurately estimate the effect of DACA participation, I expand on equation 2.6 by including a measure of the probability an observed DACA eligible no-citizen did not receive deferred action and work authorization through DACA by time.

$$P(R = \widehat{0|C, T}, E = 1) = \left(1 - \frac{(\text{Total DACA Recipients})_{ct}}{(\text{Total DACA Eligible Population})_{c}}\right) \cdot Eligible_{it} \quad (2.14)$$

Where $P(R = \widehat{0|C, T}, E = 1)$ is zero for years 2005 to 2012 and if the non-citizen is DACA-ineligible. I estimate the following equation:

$$E[Y|C, X, W, Z, S] = \beta_0 + \beta_1 P(\widehat{R = 1|C}, T, E) + \beta_2 P(R = \widehat{0|C, T}, E = 1) + \beta_3 X + \beta_4 W + \beta_5 Z + \gamma_c + \gamma_t + \gamma_s + \gamma_s t \quad (2.15)$$

all other controls are the same as in equation 2.6. Standard errors are clustered at the state-year level. The control group is now composed of only DACA-ineligible non-citizens.

The first coefficient of interest is β_1. This corresponds to the effect of DACA on DACA recipients relative to ineligible non-citizens. The coefficient β_1 is analogous to β^* in equation 2.6 except that the control group is composed of only ineligible non-citizens. If there are no spillover effects then β_1 will equal the estimates of β^* in Section 2.5.

The second coefficient of interest, β_2, corresponds to the effect of DACA on observed DACA-eligible non-participants relative to ineligible non-citizens. When discussing the spillover effects of DACA on observed eligible non-participants, I am including authorized immigrants that meet the observed DACA eligiblity requirements (age and education). As the eligiblity indicator is heavily contaminated with authorized immigrants, the coefficient β_2 will be a lower bound on the spillover effect of DACA on actual DACA-eligible non-participants.

Prior to discussing the results, it is of value to discuss what might be causing the surprisingly low participation rates (given the perceived benefits) in the program. As stated above, an estimated 63% of the immediately eligible population and roughly 50% of the potentially eligible population applied and received deferred action and work authorization through the DACA program (Hipsman et al., 2016). The participation rate is surprisingly low given the high perceived benefits associated with DACA and the actual benefits estimated in this paper.

The National UnDACAmented Research Project (NURP) is the first national survey of DACA recipients (and non-participants). While not a representative sample, the NURP survey provides valuable information on the barriers that have limited participation into DACA. The NURP survey is a national survey of 2,684 DACA-eligible individuals between the age of 18 to 32 conducted in 2013 (Gonzalez and Bautista-Chavez, 2014). Among the 2,684 respondents, 2,2440 DACA-eligible youth had been approved for DACA. The remaining 244 individuals who met the DACA requirements had not yet applied to the program. While the fee appears to be trivial compared to the benefits of receiving DACA, Gonzalez and Bautista-Chavez (2014) show more than 43% of DACA-eligible non-applicants stated that they could not afford the $465 application fee. Indicating that financial constraints are a considerable barrier for not applying. Lack of understanding of the benefits associated with DACA also appear to play a role as 30% of non-applicants stated they are waiting for better options (Gonzalez and Bautista-Chavez, 2014). Other factors documented in Gonzalez and Bautista-Chavez (2014) that prevented DACA-eligible immigrants from applying are missing paperwork (22%), legal reasons (17%), fear of sending their personal information to the government (15%), and 10% indicated they did not know how to apply.

Table 2.9 and Table 2.10 provide the estimated effects of DACA on recipients and the spillover effects of DACA on eligible non-participants for labor market outcomes and income respectively relative to non-citizens. Row 1 on both tables provides the estimated effects of DACA on recipients (β_1). Row 2 on both tables provides the estimated effects of DACA on eligible non-participants (β_2).

In relation to the preferred model, the estimated effects of DACA on the labor market outcomes, school attendance, and income of DACA recipients are similar in magnitude once separating out the spillover effects of DACA on eligible non-participants. Schooling is 1.1 p.p. larger than in the preferred model. The effect of DACA on recipients likelihood of self-employment shows a statistically significant 1.0 p.p. (20%) decrease. The biggest difference is on the effects of DACA on recipients hourly wage. The effect on hourly wage when using only DACA-ineligible non-citizens as a control group is now a statistically significant 7.5% increase.

For eligible non-participants, DACA did not have an impact in the likelihood of working. A 1.4 p.p. decrease in the probability of participating in the labor force is documented driven by unemployed leaving the labor force. School attendance among DACA eligible non-participants decreased by 2.3 p.p. DACA had significant negative spillover effects on eligible non-participants income. Relative to ineligible non-citizens DACA eligible non participants saw a 30.3% decrease in total income driven by a 20.3% decrease in wage and salary income, a 14.7% decrease in income from other sources, and a 11.7% decrease in hourly wages.

Two important notes need to be made with regards to these estimates. Second, as I can only estimate observed-eligible non-citizens this proxy is severely contaminated with authorized non-citizens which will attenuate the estimates towards zero. Second, these effects may be capturing eligible individuals attempting to become eligible but have not yet been approved at time t. The results may be driven by change in composition on who is classified as eligible non-participants over time. In the next section, I deal with this concern in an event study framework where I analyze the effects of DACA-eligible non-citizens who never participated in DACA.

As the model estimates the effect of DACA on eligible non-participants at time t, the coefficient is capturing the effect of DACA on two groups. One group are eligible non-participants at time t that will eventually apply and be approved for DACA k periods in the future. For instance, individuals who do not yet have the financial means to apply for DACA may seek employment first to save for the cost of the application process. The results may also be driven by change in composition on who is classified as eligible non-participants by time t. A second group is composed on eligible non-participants that have no intention of ever participating in DACA due to seeking other options or not meeting the other requirements (criminal history) that are not observable in the ACS data. Additionally, due to contamination in the observed DACA eligible indicator, the estimated spillover effects of DACA on the Eligible non-participating population are attenuated towards zero.

2.7 Event Study Model

Identification in the prior sections relies on the assumption that in the absence of DACA, DACA participants would have exhibited similar trends to ineligible non-citizens that did not participate in DACA. To test the plausibility of this assumption I estimate an event study model. There are two benefits of this model; (1) the parallel trends assumption can be tested by comparing the conditional trends prior to the enactment of DACA and (2) one can visualize any dynamic effects of DACA participation on labor market outcomes and income.

I construct a measure for each non-citizen in the sample of the probability of ever participating in DACA during the years 2013 to 2018. This is equivalent to assigning $P(R|C,T,E)$ at year 2018 to all years. To test the parallel trend assumption for eligible non-participants, I also construct a measure of the probability an eligible non-citizen never participated in DACA during years 2013 to 2018. This also alleviates concerns that the spillover effects in the prior section are driven by compositional changes between the participating and eligible non-participating populations.

The event study model is as follows:

$$
\begin{aligned}
E[Y|C,T,E,X,W,Z,S] = &\sum_{j\neq t^*-1} \delta_j \cdot YEAR_{j=t} \cdot P(R=1|\widehat{C,T=2018},E) \\
&+ \sum_{j\neq t^*-1} \alpha_j \cdot YEAR_{j=t} \cdot P(R=0|C,\widehat{T=2}018,E=1) \\
&+ \beta_1 P(R=1|\widehat{C,T}=2018,) + \beta_2 P(R=0|C,\widehat{T=2}018,E=1) \\
&+ +\beta_3 X + \beta_4 W + \beta_5 Z + \gamma_c + \gamma_t + \gamma_s + \gamma_s t
\end{aligned}
\tag{2.16}
$$

The variable $P(R=1|\widehat{C,T=}2018,E)$ is the probability a non-citizen was ever a DACA recipient. The variable $P(R=0|C,\widehat{T=2}018,E=1)$ is the probability an eligible non-citizen never participated in DACA during the post DACA period. Each measure is interacted with year dummies. The year 2012 is the reference year and omitted. The control group are ineligible non-citizens. All other controls are the same as Equation 2.6. Standard errors are clustered at the state-year level.

Column (a) in Figure 2.4 and Figure 2.5 shows the event study estimates of the effect of DACA participation on labor market outcomes and schooling. The figures show that there was no pre-existing trends between DACA recipients and DACA ineligible non-citizens prior to 2013 for labor force participation, working, unemployment, and school attendance. Self-employment shows differential pre-trends, though, the pre-trend is counter to the effects of DACA on self-employment. Accounting for potential differential linear pre-trends in the outcome by participation status will indicate DACA participation led to significant decrease in self-employment rates. Some noticeable differences from the estimates in the alternative model are seen in unemployment which is no longer significant. The effects of wages are negative and insignificant relative to 2012 compared to the 7.5% increase documented in the alternative model.

Column (a) in Figure 2.6 and Figure 2.7 shows the event study estimates of the effect of DACA participation on income. The coefficients in the year prior to DACA are insignificant except for year 2010 in the effects of total income and wage and salary income.

Immediately after the enactment of DACA, large effects are estimated for nearly all variables for DACA participants. After 2015, the effects have been decreasing with sharp drops in 2017 and 2018. A number reasons can be attributed for the decrease in the estimates labor market effects of DACA after 2015. First, the effects of DACA on school show a delayed response with a rapid rise in school attendance after 2015. This indicates a shift from the labor market to schooling. Second, almost 15% of DACA participants have had their status terminated or have let their status expire. Without work authorization these undocumented immigrants can no longer legally participate in the labor market. Third is the start of the Trump administration which has attempted to fully terminate DACA since 2017. Any change in participants perceived risk or deportation or their expected time in the US will have an impact on their labor market outcomes.

Column (b) in Figures 2.4, 2.5, 2.6, and 2.7 shows the event study estimates of DACA on eligible non-participants. The pre-trends provide evidence of the validity of the parallel trend assumption. The results show eligible non-participants experienced and immediate drop unemployment rate relative to 2012 but a gradual decrease in labor force participation reaching a significant 5 p.p. reduction by 2018. Across income sources, only income from other sources show a significant impact. The results of spillover effects on total income, wage income, and hourly wage of eligible non-participants are not robust.

2.8 Robustness Checks

The estimates from the preferred model in Section 2.5 and the modified model controling for possible spillover effects on DACA-eligible non-participants in Section 2.6 are the best available estimates of the average treatment-on-the-treated effects of DACA with data that is publicly available. The results indicate that DACA has lead to significant improvements in on the labor market outcomes and the incomes of those that participated in the program relative to other non-citizens.

The fundamental assumption needed for the treatment-on-the-treated estimates produced above to be unbiased is the assumption that the variation in the share of DACA participants across country-of-origin and time only affect outcomes through the unobserved participation decision. If the constructed probability is measuring other time varying country specific unobserved variables besides the participation decision that are correlated with the outcome of interest across time, the results will be biased.

Another issue with the preferred model is that I am using a proxy variable that only takes advantage of variation across country-of-origin rather than the ideal indicator of DACA participation, there will be issues relating to mis-measurement. One source of this mis-measurement comes from using observed DACA eligiblity requirements to identify eligibility which misses the non-citizens' unobserved undocumented status and their unobserved criminal history. This means I am assigning a positive probability to documented immigrants and ineligible undocumented immigrants that meet the observed education and age requirements. The country-of-origin proxy also misses important demographic and geographic variation among DACA participants that would allow for finer estimation of the average treatment-on-the-treated effects. There is also mis-measurement in the administrative data on DACA. In the data some participants have demographic characteristics that fail the requirements for eligibility or are missing (USCIS, 2018). In 2018, the USCIS did not have country-of-origin data for 2,100 DACA recipients recipients (USCIS, 2018).

To strengthen the validity of the estimates presented so far, I perform a number of robustness checks. First, I take advantage of state-level variation in the share of DACA participation among DACA eligible non-citizens. Second, I perform a placebo test of naturalized immigrants. Third, I perform an alternative approach suggested by Lubotsky and Wittenberg (2006) that takes advantage of all available sources of variation available simultaneously.

2.8.1 Variation in DACA Participation Across State-of-Residence

The preferred model uses variation in the probability an observed DACA-eligible non-citizen is a DACA recipient across country-of-birth and time. I create a similar measure using variation across state-of-residence. This variable is defined in equation 2.8. Using this alternative measure is less preferred than using country-of-variation. First, there is considerably less variation with 50 states (plus DC). Second the model controls for state fixed effects and state time trends. This will cause the model absorb considerable variation in this instrument.

Even with these drawbacks, estimating the model using the alternative measure will provide a valuable robustness checks on the results of the preferred model. I take advantage of this other source of variation by estimating equation 2.6 and replacing the country-of-origin measure with the alternative state-of-residence measure.

$$E[Y|C,X,W,Z,S] = \beta_0 + \beta^* \widehat{P(R=1|S,T,E)} + \beta_2 X + \beta_3 W + \beta_4 Z + \gamma_c + \gamma_t + \gamma_s + \gamma_s t \quad (2.17)$$

where $\widehat{P(R=1|S,T,E)}$ is the probability of a DACA eligible non-citizen being a DACA recipient at time t using variation across state s. All other controls are the same as with the preferred equation 2.6. Errors are clustered at the state-year level.

As with the preferred model, I make the key assumption that the constructed probability only affects outcome y through the unobserved DACA participation, or that

$$\widehat{P(R=1|S,T,E)} = P(R=1|C,T,E,X,W,Z,S) = E[R|C,T,E,X,W,Z,S] \quad (2.18)$$

If the assumption made in both equation 2.5 and equation 2.18 hold, both models should provide similar average treatment-on-the-treated effects of DACA.

Row 1 of Table A.2 shows the effect of DACA on labor market outcomes using variation across state-of-residence. Row 1 of Table A.3 shows the effect of DACA on income using variation across state-of-residence. As would be expected from the state fixed effects and the state time trends absorbing considerable variation in the variable of interest, the standard errors are considerably larger than in the preferred model. Even with the larger standard errors the estimated effects are still statistically significant as in the preferred model. The magnitude of the effects of DACA on recipients likelihood of working are similar. The magnitude on the effect of DACA on recipients labor force participation, usual hours worked, total income, income from wages and are smaller than in the preferred model. The magnitude of the effects on the likelihood of being unemployed and income from other sources are larger magnitude than in the preferred model.

The most noticeable difference are found in the effect of income from other sources, school attendance, and wages. Using the alternative proxy produces and effect on income from other sources that is negative at a 19% compared to a positive 7% increase in the preferred model. The effect on schooling is statistically and economically insignificant compared to a significant 2 p.p. decrease. Lastly, the effect of DACA on recipients wages is now negative when using state variation.

In nearly all outcomes of interest, there is a downward change in the estimates using the state-of-residence measure compared to the preferred model using country-of-origin measure. As the alternative measure uses significantly less variation, considerably more non-citizens in the sample will be mis-measured which will attenuate the results towards zero. For instance, an observed eligible immigrant from England living in California will have a constructed probability of 0.491 rather than the actual probability of 0. This though, will not explain why some estimates are more negative then the preferred model.

2.8.2 Placebo Test on Naturalized Citizens

A major concern in the estimates produced in the preferred model is in the underlying assumption that there are no country-year unobservable covariates that are correlated with the constructed measure and outcomes interest. For the case of the measure using state variation, that there are no state-year unobservable covariates correlated with the constructed measure and outcomes of interest. In other words, that the measures are capturing only the participation decision and nothing else. I perform a placebo test on naturalized citizens to test this assumption. If there are other unobserved time varying factors at the country-level or state-level that are being captured by the constructed measures, then the coefficient on the DACA recipient measures using a sample of naturalized citizens citizens should be similar.

I estimate equation 2.6 and 2.17 on a sample of naturalized citizens ages 18-35 with a high school degree or more. I treat all naturalized citizens as non-citizens and perform the same procedure to assign a probability to to each. Table A.4 shows the placebo test for the labor market outcomes while Table A.5 shows the placebo test for income outcomes. For each Table, Row 1 uses the country-of-origin variation measure as in the preferred model. Row 2 uses the state-of-residence variation measure.

The placebo results are consistent with the key assumption being satisfied. Only 4 of the 22 regressions produced a marginally statistically significant effect. For 3 of the 4 significant coefficients, the magnitude of the effects are a fifth or half the size compared to the preferred model. Only one coefficient, on the effect of DACA on hourly wages using state-of-residence measure, has a similar magnitude.

2.8.3 Lubotsky and Wittenberg (2006) Approach

A disadvantage of the prior models estimates is that they are using each source of variation separately. I next implement an interpretation approach suggested by Lubotsky and Wittenberg (2006) (henceforth L-W). The L-W method includes all variables that proxy for the unobserved variable of interest in the estimating question. I estimate equation 2.6 but include the two constructed probability measures.

$$\begin{aligned} E[Y|C,X,W,Z,S] = \beta_0 + \beta_c^* \widehat{P(R=1|C,T,E)} + \beta_s^* \widehat{P(R=1|S,T,E)} \\ + \beta_2 X + \beta_3 W + \beta_4 Z + \gamma_c + \gamma_t + \gamma_s + \gamma_s t \quad (2.19) \end{aligned}$$

L-W show that the weighted sum of the coefficients of the multiple proxy variables that measures the same unobserved variable of interest minimizes the attenuation bias from the mis-measured proxies. The attenuation minimizing estimate of β^* is written as

$$b^\rho = \rho_c \cdot \beta_c^* + \rho_s \cdot \beta_s^* \quad (2.20)$$

Where b^ρ is the minimizing attenuation bias estimate of β^*. The weight of each proxy ρ_j must be estimated. The estimate of ρ_j is $\hat{\rho}_j = \frac{cov(y,P(recipient)_j)}{cov(y,P(recipient)_1)}$. The weights are to ensure all variables are scaled the same and the weights must be scaled so that one weight is equal to one. The weight for the probability measure using country-of-origin variation is chosen as the base ($\rho_c = 1$) so that they can be compared to the estimates from Table 2.3 and Table 2.4.

Under the assumption that both variables only capture the underlying unobserved variable (recipient), the estimated effects for all outcomes of interest in this specification will be larger than in the preferred model using the mis-measured constructed probability the greater the measurement error. Row 1 of Table A.6 shows the estimates of the L-W approach for labor market outcomes. Row 1 of Table A.7 shows the estimates from the L-W method for the income outcomes. In the L-W model There is no significant difference between the coefficients produced in the preferred model using variation across country-of-origin and taking advantage of both sources of variation using the L-W method. Besides for estimates in the effect of DACA on recipient's income from other sources and usual hours of work, the magnitude in the L-W estimates are slightly lower. Nearly all the coefficients are larger than the estimates from the preferred model using variation across state-of-residence. This would be consistent with state-of-residence recipient measure having more measurement error. The estimate of the effect of DACA on wages using both measures is of opposite sign and statistically insignificant.

The effects on school attendance is now positive and a large but statistically insignificant 22.3 p.p. compared to a 34 p.p. decrease in the preferred model. This is an unrealistic increase. One possible explanation for this is the fact the weights have to be estimated. As the purpose of the weights is for scaling and given the design of the two variables capture the probability an individual is a DACA recipient, an assumption can be made that the weights should equal one. In fact, this will come from the assumptions made to derive assumption 2.5 and assumption 2.18.

Row 2 of Table A.6 and Row 2 of Table A.7 shows the estimates of the the weight $\hat{\rho_s} = \frac{cov(y,P(recipient)_s)}{cov(y,P(recipient)_c)}$.[12] The range of values in the weights vary considerably. In the case of unemployment, the weight is negative. For schooling, the weight is a large 2.2 and negative. This is over twice as large as the other weights and of opposite sign. Aside from concerns in the estimation of the weight, the variables may also be capturing additional unobservable covariates aside from individuals DACA recipient status.

Rather than estimate the weight, I next implement the L-W method assuming ρ is equal to one as implied by assumption 2.5 and assumption 2.18. Row 2 of Table A.6 shows the estimates of the L-W approach for labor market outcomes while Row 2 of Table A.7 shows the estimates from the L-W method for the income outcomes when ρ is assumed to be equal to one. The estimates are larger for likelihood of working, likelihood of of being unemployed, and income from other sources. The estimated effects of other outcomes are slightly lower when using $\rho = 1$ rather than the estimated ρ. Notably, the effect on schooling is a statistically insignificant 1.2 p.p. increase. While the effect on wages are a significant 10.4% decrease which is similar to that when using state-of-residence measure.

[12]The weight $\hat{\rho_c}$ is set equal to 1 as it is the reference weight used for the scaling.

2.9 Concluding Remarks

I construct a novel measure of DACA participation using publicly available data from the USCIS to estimate the effect of DACA on recipients. This allows for the first estimation of the treatment-on-the-treated effects of the DACA program. These are the best estimates available with currently available administrative data. I find that the reduction of labor market frictions through deferred action and work authorization from participating in the DACA program increased the likelihood of working by 11.3 percentage points (p.p.) or 17.1%. This is driven by a 9.6 p.p. increase in labor force participation and a 3.0 p.p. decrease in unemployment among DACA recipients. I find statistically significant decrease in the likelihood of a DACA recipient being self-employed with a magnitude that is economically meaningful at 18%. As self-employment is a proxy for informal employment among the undocumented population, this implies economically significant movement from the informal to the formal labor market after participating in DACA. Unlike previous studies on DACA, I find positive effects on school attendance on DACA recipients. The effects are delayed with increases occurring after 2015.

DACA recipients saw an increase in total income of 108.1%, equivalent to a $15,510 increase in total income from their pre-DACA levels. The increase in total income is driven entirely by increases in wage and salary income. With a total of 824,000 participants since its enactment in 2012, DACA moved 101,000 to 103,000 undocumented youth into employment. The effect on income implies a $12.8 billion increase in total income for the entire DACA participating population.

Similar results are obtained when using alternative measure constructed from variation in the participation rate among DACA-eligible non-citizens across state-of-residence. The results are also robust to using an alternative method, the L-W method, that takes advantage of all variation from both proxies. Although the effects on wages and education are not robust to the primary measure using variation across country-of-origin.

The scaling factor between the estimated treatment-on-the-treated effects and intent-to-treat effects in this paper ranges from 1.8 to 7.9 times larger depending on the outcome of interest. This shows the treatment effects on recipients of deferred action can not simply be extrapolated from the effects on those eligible using the standard assumptions. While self-selection is a likely source of the significant variation in the scaling factor, this paper can not exclude the possibility the differences are driven by heterogeneous effects of DACA. I find significant heterogeneity in the treatment-on-the-treated effects across region-of-origin. I find that recipients from Asia had the largest labor market benefit from DACA. Latin Americans saw significantly lower labor market benefits compared to Mexican recipients but had the largest increase in school attendance from receiving deferred action and work authorization. As non-Mexicans become a larger share of the undocumented population, understanding these heterogeneous effects are important to estimate the cost and benefits of future amnesty programs.

I also document negative spillover effects on Eligible non-participants. I find eligible non-participants had 1.4 p.p. decrease in labor force participation, driven by those in unemployment. Eligible non-participants also reduced their likelihood of attending school after DACA is announced. Policy makers should carefully understand why certain eligible undocumented immigrants are not participating in the program to better designed amnesty proposals that meet their objectives. Policy makers must also take into account spillover effects on this group to see how the costs and benefits of the proposed programs are altered.

Left for future research is the puzzle in understanding why a significant portion of the estimated DACA eligible population did not participate in the program. As shown by the literature and in this paper, large benefits are left in the table among non-participants. Future research should investigate the differences in participation rates across state-of-residence and country-of-origin. This will provide policy markers information needed to increase participation rates in future immigration reform programs.

2.10 Chapter 2 Tables

Table 2.1: Descriptive Statistics of Probability Measure by Country-of-Origin (2018)

Country of Origin	(1) Eligible Population	(2) Total Approvals	(3) Constructed Probability
Mexico	906,643	643,372	0.710
El Salvador	61,697	29,459	0.477
Guatemala	42,562	20,721	0.487
Honduras	29,776	18,962	0.637
Peru	21,526	9,249	0.430
Korea	41,638	8,967	0.215
Brazil	17,497	7,555	0.432
Ecuador	17,789	6,842	0.385
Colombia	30,475	6,710	0.220
Argentina	8,957	4,917	0.549
Philippines	4,3801	4,766	0.109
Jamaica	27,908	3,505	0.126
India	27,523	3,242	0.118
Dominican Republic	38,775	3,240	0.084
Venezuela	14,020	3,172	0.226
Trinidad and Tobago	13,118	2,620	0.200
Uruguay	3,740	2,464	0.659
Bolivia	4,260	2,107	0.495
Costa Rica	5,248	2,093	0.399
Poland	14,253	1,852	0.130
Chile	4,524	1,784	0.394
Pakistan	8,898	1,720	0.193
Nicaragua	14,458	1,645	0.114
Nigeria	6,626	1,320	0.199
Guyana/British Guiana	8,416	1,293	0.154
Belize/British Honduras	2,856	1,021	0.358
Canada	30,571	990	0.032
China	28,986	982	0.034
Kenya	5,063	874	0.173
Indonesia	3,479	836	0.241

Notes: Authors own calculations for the constructed probability in the year 2018. Eligible population is calculated using ACS data from 2013 to 2018. Total approvals are from 2018 USCIS publicly available data (USCIS, 2018). Constructed Probability measure is the ratio between total approvals over eligible population by country.

Table 2.2: Pre-DACA Characteristics of Participants and Comparison Groups

Variables	(1) Recipients	(2) Eligible Non-Participants	(3) DACA Ineligible	(4) Full Sample
Individual Characteristics				
Age	23.9	23.8	28.6	27.6
Years in US	15.7	15.2	6.4	8.3
Age enter US	8.2	8.7	22.3	19.3
In School	25.1	37.2	21.1	23.1
HS Degree	59.5	43.4	38.2	41.4
Some College	35.0	43.5	25.3	28.3
College Degree	5.6	13.1	36.5	30.3
Male	53.1	52.6	52.0	52.4
Married	27.9	21.7	51.8	46.0
White	2.6	17.4	17.4	16.2
Black	1.9	13.6	8.2	8.4
Asian	3.3	22.7	29.1	25.5
Hispanic	91.6	44.3	43.2	48.2
Born in Mexico	77.7	23.1	27.5	32.0
Born in Europe	1.0	12.4	10.5	10.0
Born in Asia	3.5	23.9	32.3	28.0
Born in Latin America	17.2	31.5	21.2	22.2
C. Outcomes				
Labor Force Participation	74.7	72.9	72.5	72.3
Working	66.2	63.9	67.0	66.8
Unemployed	11.3	12.4	7.6	8.4
Usual Hours Worked	27.9	26.7	29.0	28.0
Self-Employed	5.0	4.5	6.6	6.1
Total Income	$15,117	$16,255	$23,929	$22,132
Wage Income	$14,165	$15,155	$22,309	$20,649
Other Income	$245	$329	$406	$374
Welfare Income	$40	$35	$31	$32
D. Identifiers				
Observed Eligible	100.0	100.0	0	23.5
Probability Recipient	57.2	27.7	0	8.6

Notes: The sample for the summary statistics includes non-citizens who are ages 18–35 and have at least a high school degree in the years 2005 to 2012. All binary variables are represented in percent (%) terms. All values represent the weighted sum using ACS weights times the respective weight. The weights used for each column are as follows; (1)person weight times probability DACA recipient, (2) person weight times probability eligible non-participant, (3) person weight times indicator DACA ineligible, and (4) person weight.

Table 2.3: The Effects of DACA on Labor Market Outcomes

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(6) School
DACA Recipient	0.096***	0.113***	-0.030***	3.802***	-0.009*	0.034**
	(0.012)	(0.012)	(0.006)	(0.591)	(0.004)	(0.012)
Pre-DACA Mean	0.747	0.662	0.113	27.929	0.050	0.251
N	618,450	618,450	432,284	618,450	498,081	618,450
R^2	0.152	0.146	0.034	0.201	0.029	0.333

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.4: The Effects of DACA on Income (IHS)

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
DACA Recipient	1.026***	1.081***	-0.015	-0.023	0.016
	(0.131)	(0.132)	(0.036)	(0.013)	(0.024)
Pre-DACA Mean	\$15,117	\$14,165	\$245	\$40	\$12.63
N	618,450	618,450	618,450	618,450	423,477
R^2	0.181	0.159	0.026	0.014	0.300

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. *Wage* - inverse hyperbolic sine (IHS) transformation of constructed wage. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.5: The Effects of DACA-Eligibility on Labor Market Outcomes

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(6) School
eligible x Post	0.019***	0.030***	-0.016***	0.671**	-0.003	0.008
	(0.005)	(0.005)	(0.003)	(0.261)	(0.002)	(0.005)
Recipient/Eligible	5.05	3.77	1.88	5.66	3.00	4.25
N	618,450	618,450	432,284	618,450	498,081	618,450
R^2	0.130	0.129	0.031	0.183	0.023	0.334

Standard errors in parenthesis

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *Income (IHS)* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for Post-DACA implementation dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.6: The Effects of DACA-Eligibility on Income (IHS)

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
eligible x Post	0.127**	0.193***	-0.075***	-0.010	- 0.061***
	(0.055)	(0.057)	(0.016)	(0.006)	(0.011)
Recipient/Eligible	7.87	5.60	0.13	2.30	-0.26
N	618,450	618,450	618,450	618,450	423,477
R^2	0.168	0.139	0.011	0.010	0.30

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS (Panel A) and 2005-2016 for Panel B. All regressions control for Post-DACA implementation dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.7: Heterogeneity in the Effects of DACA on Labor Market Outcomes

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(5) School
$P(R\|S,T)$	0.087***	0.111***	-0.038***	3.202***	-0.006	0.007
	(0.013)	(0.014)	(0.007)	(0.680)	(0.006)	(0.013)
$P(R\|S,T)$*latin	-0.026*	-0.002	-0.029***	-1.222*	-0.048***	0.054***
	(0.012)	(0.013)	(0.009)	(0.511)	(0.008)	(0.009)
$P(R\|S,T)$*asia	0.073*	0.066	0.032	0.315	0.036	0.123***
	(0.036)	(0.036)	(0.033)	(1.393)	(0.021)	(0.033)
$P(R\|S,T)$*euro	-0.009	-0.010	-0.002	-9.374***	-0.260***	0.098
	(0.062)	(0.072)	(0.051)	(2.568)	(0.046)	(0.059)
$P(R\|S,T)$*rest	-0.410***	-0.477***	0.172*	-24.451***	0.044	-0.093
	(0.091)	(0.099)	(0.083)	(3.880)	(0.045)	(0.076)
N	618,450	618,450	432,284	618,450	498,081	618,450
R^2	0.152	0.146	0.034	0.201	0.030	0.333

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.8: Heterogeneity in the Effects of DACA on Income (IHS)

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
$P(R\|S,T)$	0.764***	0.880***	-0.140***	-0.019	-0.110***
	(0.147)	(0.142)	(0.041)	(0.015)	(0.027)
$P(R\|S,T)$*latin	-0.183	0.208	-0.135**	0.007	0.086***
	(0.116)	(0.133)	(0.041)	(0.016)	(0.017)
$P(R\|S,T)$*asia	-0.260	0.406	-1.226***	-0.029	-0.657***
	(0.356)	(0.368)	(0.152)	(0.042)	(0.072)
$P(R\|S,T)$*euro	-0.914	0.551	-0.552*	0.038	-0.705***
	(0.650)	(0.768)	(0.238)	(0.128)	(0.112)
$P(R\|S,T)$*rest	-6.017***	-6.145***	-0.246	0.161	-0.762***
	(0.948)	(1.030)	(0.347)	(0.138)	(0.143)
N	618,450	618,450	618,450	618,450	423,477
R^2	0.181	0.158	0.026	0.014	0.300

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.9: The Spillover Effects of DACA on Labor Market Outcomes

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(6) School
DACA Recipient	0.103***	0.115***	-0.024***	4.250***	-0.010*	0.045***
	(0.012)	(0.013)	(0.006)	(0.606)	(0.005)	(0.012)
Eligible Non-Participant	-0.014*	-0.004	-0.011**	-0.869**	0.003	-0.023***
	(0.007)	(0.007)	(0.004)	(0.284)	(0.004)	(0.005)
N	618,450	618,450	432,284	618,450	498,081	618,450
R^2	0.152	0.146	0.034	0.201	0.029	0.334

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table 2.10: The Spillover Effects of DACA on Income (IHS)

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
DACA Recipients	1.182***	1.185***	0.061	-0.025	0.075**
	(0.138)	(0.143)	(0.036)	(0.014)	(0.025)
Eligible Non-Participants	-0.303***	-0.203**	-0.147***	0.003	-0.117***
	(0.065)	(0.070)	(0.022)	(0.010)	(0.012)
N	618,450	618,450	618,450	618,450	423,477
R^2	0.181	0.159	0.026	0.014	0.300

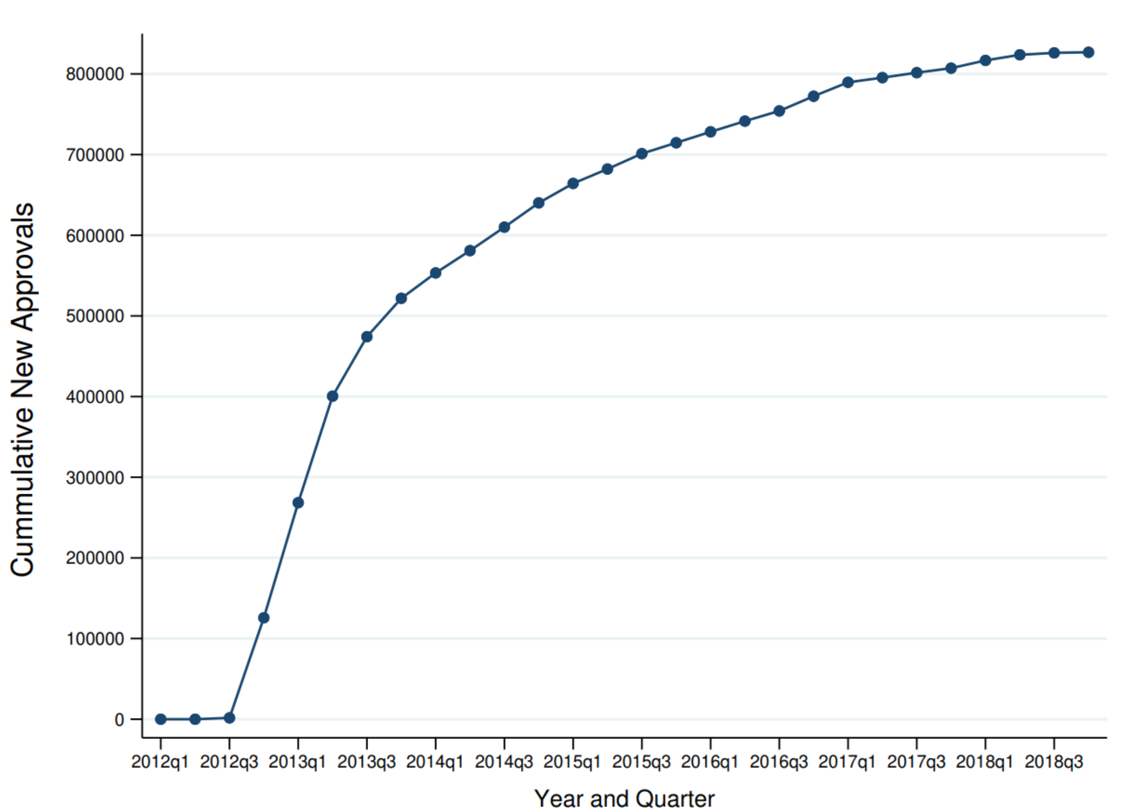
Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. *Wage* - inverse hyperbolic sine (IHS) transformation of constructed wage. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS (Panel A) and 2005-2016 for Panel B. All regressions control for Post-DACA implementation dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

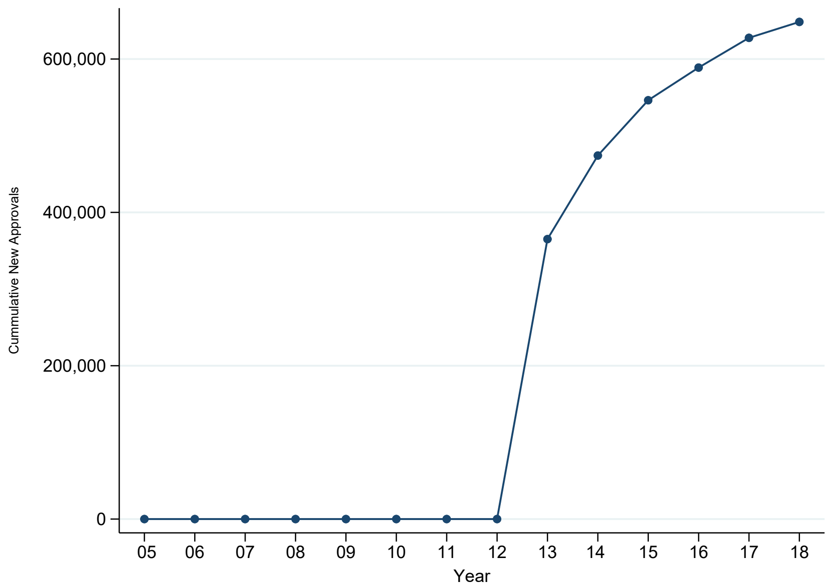
2.11 Chapter 2 Figures

Figure 2.1: Cumulative number of Initial Applications Approved by Quarter

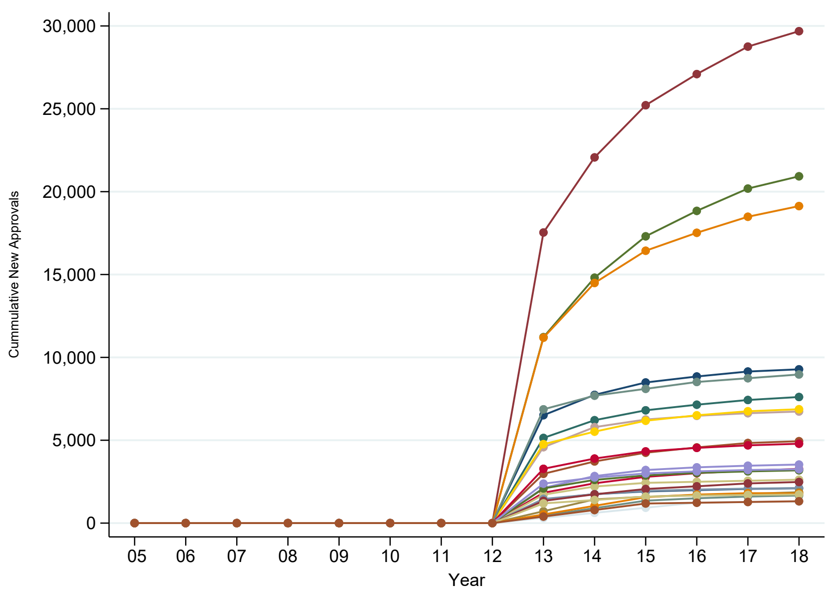


Notes: This figure shows cumulative new approvals number approved in each quarter through 2018 quarter 3. Data comes from publicly available quarterly reports published by the USCIS from 2013 to 2018 (USCIS, 2017b).

Figure 2.2: Cumulative number of Initial Applications Approved by Country-of-Origin



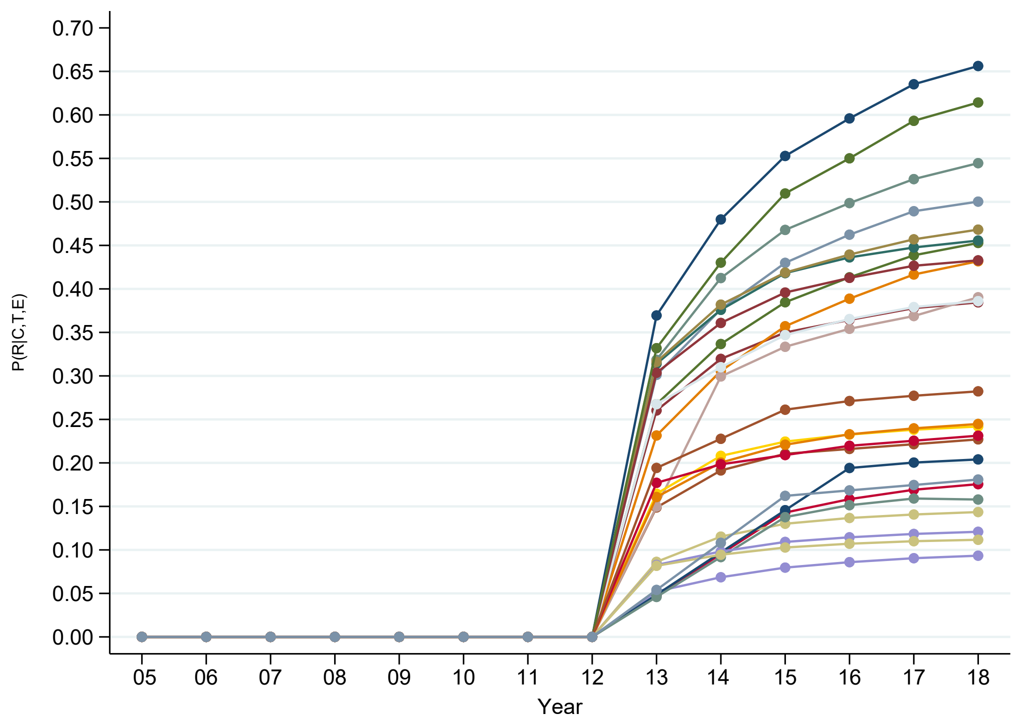
(A) Mexico



(B) Top 25 - Excluding Mexico

Notes: This figure shows cumulative new approvals number approved in each year through 2018 for the top 25 countries. Data comes from publicly available quarterly reports published by the USCIS from 2013 to 2018 (USCIS, 2017b, 2018).

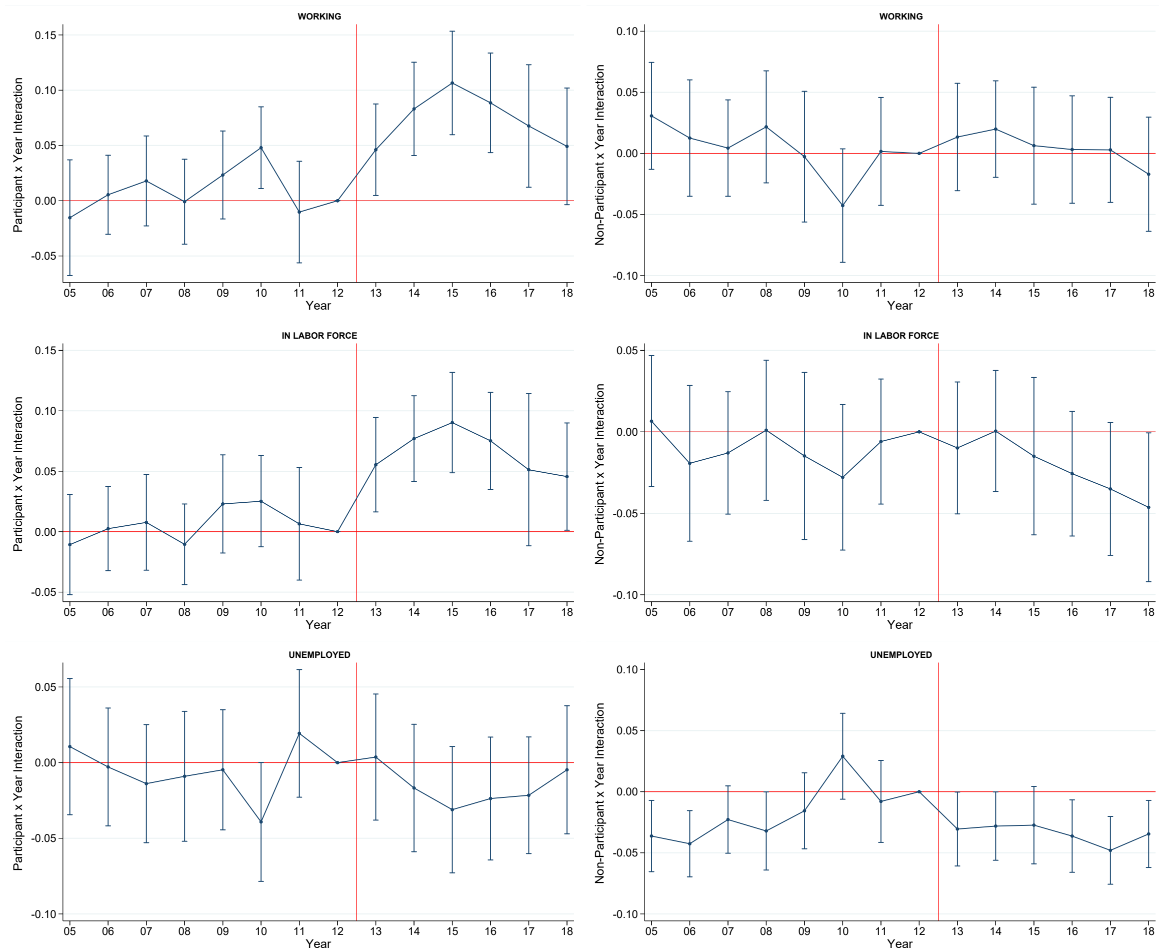
Figure 2.3: Variation in DACA Participation Measure



(A) Top 25 Countries

Notes: This figure shows cumulative probabily an observed DACA eligible immigrant is a DACA recipient in each year through 2018 for the top 25 countries. Data comes from publicly available quarterly reports published by the USCIS from 2013 to 2018 (USCIS, 2017b, 2018).

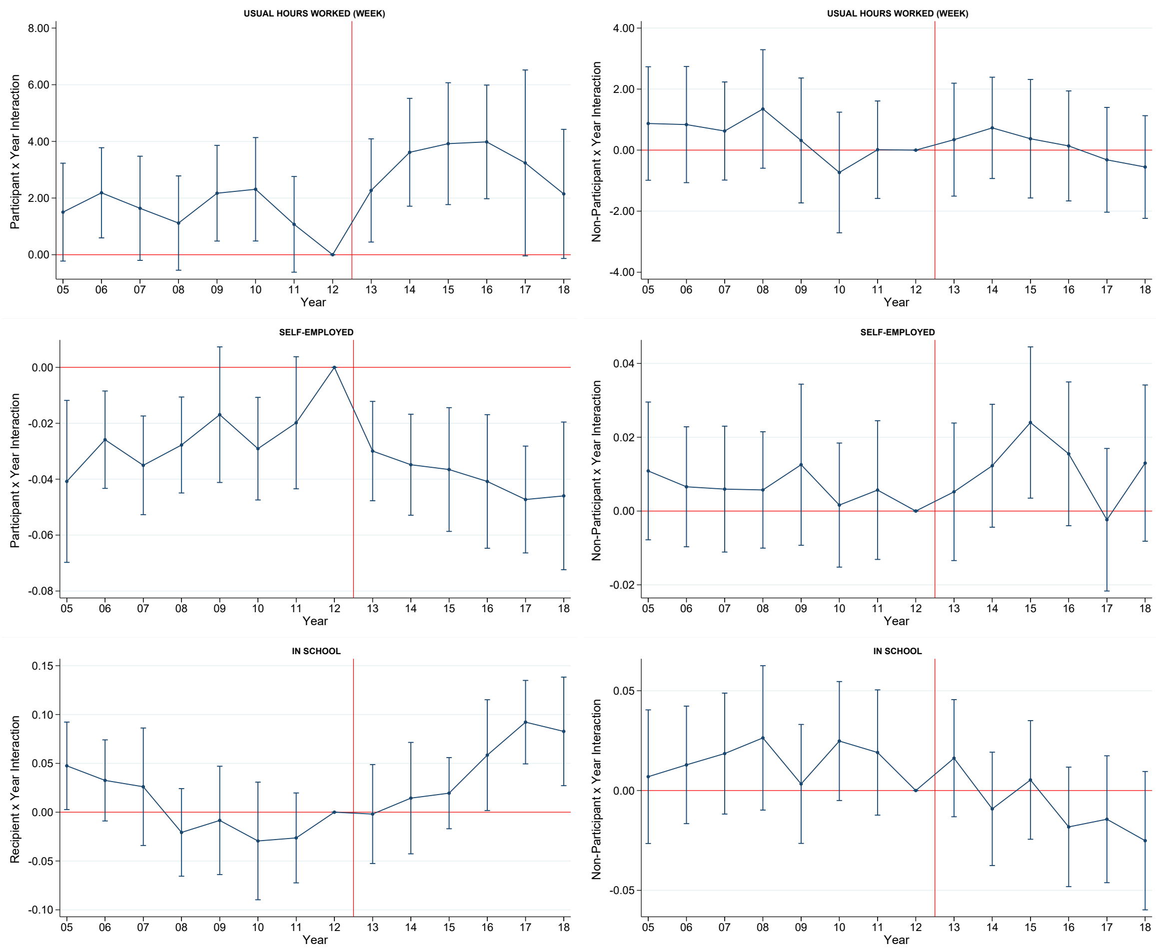
Figure 2.4: Event Study: DACA on Labor Market Outcomes I



(A) DACA Recipient (B) Eligible Non-Participant

Notes: The figure on the first column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 1|C, T, E)$ interacted with a binary variable for each year (2012 is the omitted interaction). The figure on the second column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 0|C, T, E = 1)$ interacted with a binary variable for each year (2012 is the omitted interaction). Following dependent variables were used in the regressions (up-to-down, starting with the first row): *In Labor Force* - binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Figure 2.5: Event Study: DACA on Labor Market Outcomes II

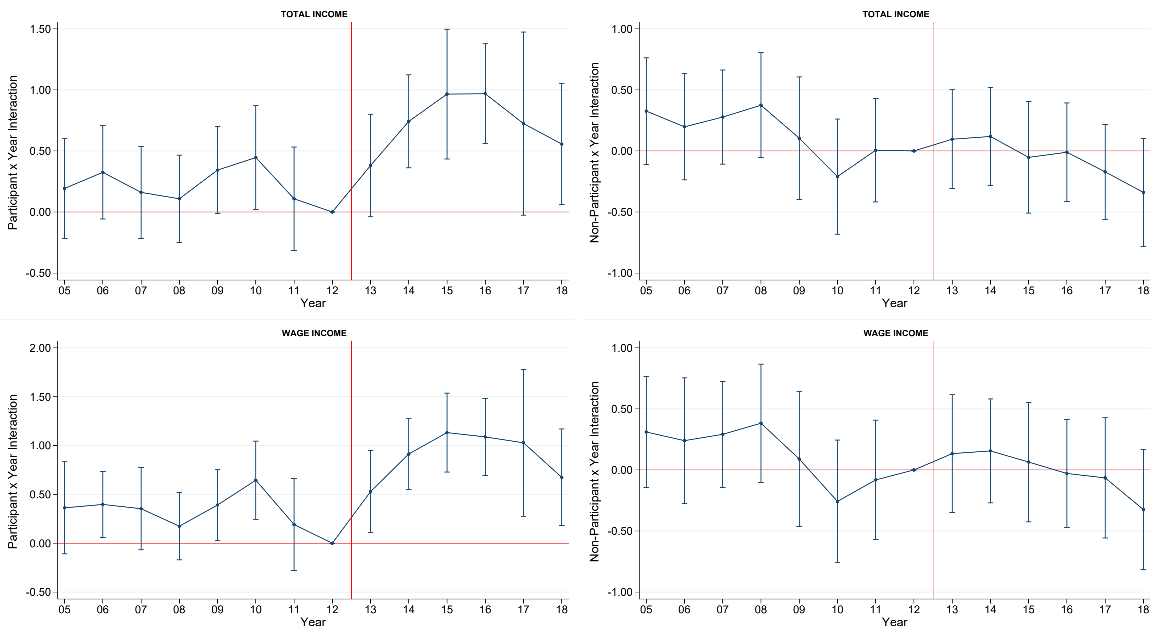


(A) DACA Recipient

(B) Eligible Non-Participant

Notes: The figure on the first column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 1|C, T, E)$ interacted with a binary variable for each year (2012 is the omitted interaction). The figure on the second column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 0|C, T, E = 1)$ interacted with a binary variable for each year (2012 is the omitted interaction). Following dependent variables were used in the regressions (up-to-down, starting with the first row): *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var equal 1 if individual is currently self-employed; *School* - binary var equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

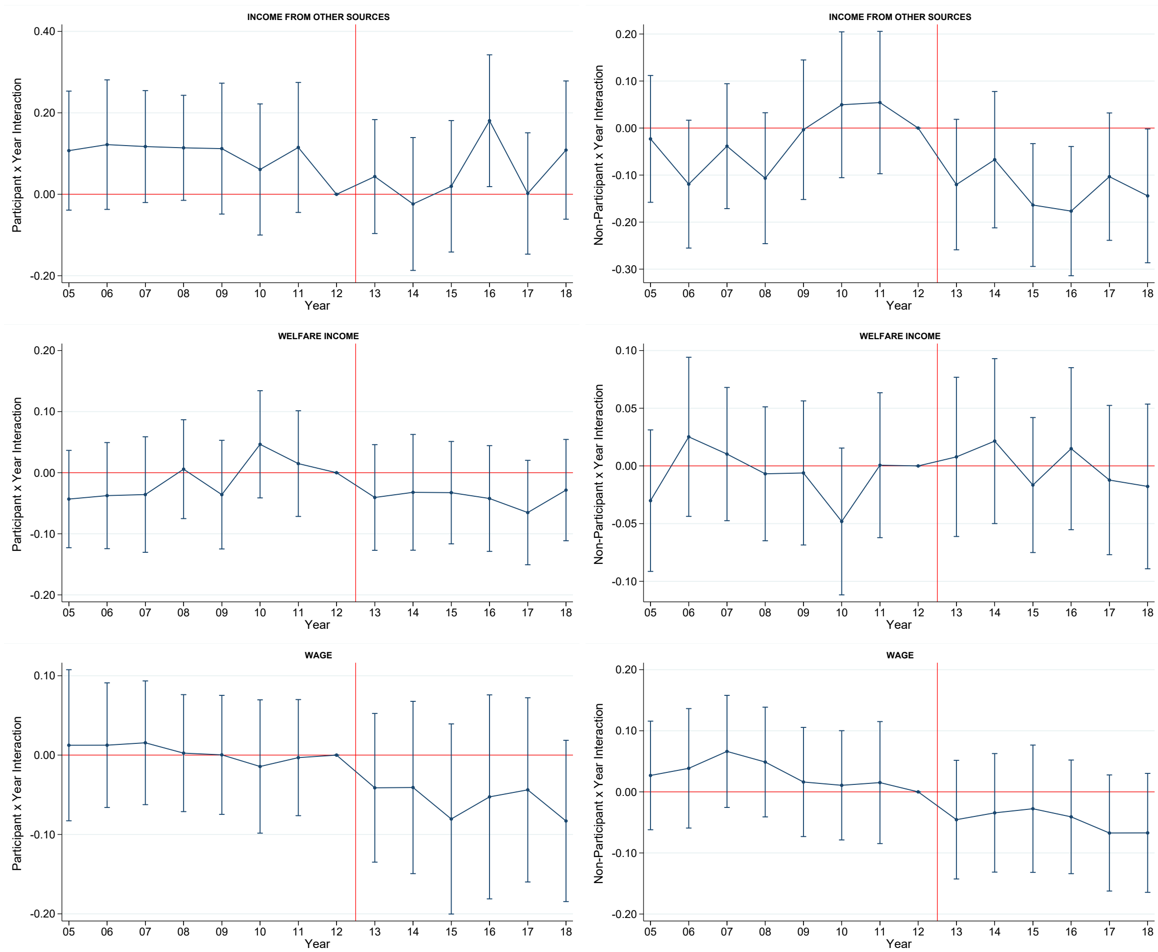
Figure 2.6: Event Study: DACA on Income I



(A) DACA Recipient (B) Eligible Non-Participant

Notes: The figure on the first column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 1|C, T, E)$ interacted with a binary variable for each year (2012 is the omitted interaction). The figure on the second column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 0|C, T, E = 1)$ interacted with a binary variable for each year (2012 is the omitted interaction). Following dependent variables were used in the regressions (up-to-down, starting with the first row): *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Figure 2.7: Event Study: DACA on Income II



(A) DACA Recipient (B) Eligible Non-Participant

Notes: The figure on the first column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 1|C, T, E)$ interacted with a binary variable for each year (2012 is the omitted interaction). The figure on the second column plots the coefficients obtained estimating Equation 1.16 with the variable $P(R = 0|C, T, E)$ interacted with a binary variable for each year (2012 is the omitted interaction). Following dependent variables were used in the regressions (up-to-down, starting with the first row): *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months; *Wages* - inverse hyperbolic sine (IHS) transformation of individuals constructed wage. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Chapter 3 Credible Interval Estimates of the Size and Legal Composition of the US Foreign-Born Population

3.1 Introduction

It has become widely accepted that, as of 2019, there are 44.9 million foreign-born individuals residing in the US of which 21.7 million are non-citizens.[1] Among the non-citizen population, approximately 11 million are estimated to be undocumented immigrants (Passel and Cohn, 2019; Warren, 2020; Baker, 2021). The estimates of the size of the undocumented population are typically reported as point estimates and usually treated as if they had *incredible certitude* (Manski, 2011) by the media and policy analysts alike. However, because documentation status must be inferred rather than directly measured, these estimates may be sensitive to underlying assumptions. As these estimates predominately use survey data an important underlying assumption is on how to account for nonsampling errors.

One type of nonsampling error that can effect the accuracy of the size and legal composition of the US foreign-born population is caused by survey nonresponse. This can take the form of unit or item nonresponse. The standard, but untested, assumption to deal with survey nonresponse which allow for point estimates to be produced is to assume that nonresponse is conditionally random. That is, conditional on a set of observable covariates, the distribution of legal status among non-respondents is the same as that of respondents. This assumption, often referred to as missing at random (MAR), is implemented as weights for unit nonresponse and imputations for item nonresponse.

The assumption of MAR will fail if nonresponse is a function of characteristics (observed and/or unobserved) not used as part of the matching criteria in the imputation or weighting procedures. Failure of this assumption is more likely with sensitive questions such as asking about citizenship status as certain groups may be hesitant to provide information on their legal status. For instance, undocumented immigrants may be more hesitant than other groups to provide information or to participate in government administered surveys due

[1] These are the official statistics produced by the Census using the American Community Survey.

to concerns about the data being used for enforcement.[2] This will lead to a failure of the the assumption that nonresponse is MAR and will bias estimates of the size and legal composition of the foreign-born population.[3]

Researchers working in the Census Bureau have cast serious doubt on the validity of this assumption being satisfied for the citizenship question in the American Community Survey (ACS) (Brown et al., 2018). Brown et al. (2018) are able to link the ACS to administrative records that contain citizenship status. Individuals who are identified as being non-citizen in administrative records are significantly less likely to respond to the citizenship question than individuals identified in the administrative records as citizens (naturalized or native-born). They also show Census tracts with noncitizen shares in the top decile have higher levels of unit nonresponse than the Census tracts with noncitizen shares at the bottom decile. Over time, the rate of unit nonresponse has also increased more rapidly among the top decile tracts compared to the bottom decile tracts (Brown et al., 2018). This suggests the imputation procedure used by the Census likely underestimates the size of the non-citizen population in the US and in turn leads to an underestimation of the size of the undocumented population.

In this paper, I use the ACS to estimate credible interval estimates of the size of the foreign-born, non-citizen, and the undocumented immigrant populations that takes into account nonsampling error from item nonresponse following the approach in Manski (2016). I focus on item nonresponse in the citizenship question as this is the question used to produce official statistics by the Census and one of the key identifiers used by researchers to assign undocumented status in survey data. Over the years, nonresponse to this question has grown rapidly reaching 7.42% of the sample by 2019 from 2.06% in 2009. The rise of item nonresponse increases the possibility of substantial nonsampling error when it comes to the point estimates of the size and legal composition of the foreign-born population in the US. Particularly given that nonresponse is rising in areas with high share of non-citizens at a higher rate than areas with high share of citizens as shown in Brown et al. (2018).

[2] While Title 13, U.S.C. prohibits the use of Census data for enforcement purposes, respondents may have this concern.

[3] Assuming, reasonably, that undocumented immigrants and legal non-citizens have a higher survey nonresponse than naturalized and native-born individuals the sign of the bias will be negative (i.e. under-estimation of the size of non-citizen and undocumented population).

The Manski approach bounds the estimated parameter by assigning the extreme case values to those individuals who did not respond to the question of interest. This approach produces the maximum degree of uncertainty caused by item nonresponse in the estimates of the size of these populations. The credible interval estimates of the size of these population groups provide the benefit of not requiring the untested assumption that, conditional on a set of observables, the distribution of legal status is the same across respondents and non-respondents.

The most popular method to estimate the size of the undocumented population is the residual method developed by Warren and Passel (1987). In the simplest terms, the residual method estimates the size of the undocumented population by subtracting the estimated number of legal immigrants residing in the US from the estimated number of the total foreign-born population. This method has been expanded to attempt to identify undocumented immigrants at the individual level in survey data. The most prominent government and public policy institutions that use the residual method have estimated size of the undocumented populated to be around 10.5 to 11.5 million people for the years 2017/2018. Passel and Cohn (2019) estimate the undocumented population at 10.5 million in 2017 while Baker (2021) at the Department of Homeland Security (DHS) estimates that 11.4 million undocumented immigrants were living in the US in 2018. Warren (2020) estimates that the undocumented population in the US was 10.5 million in 2018. Differences in the estimates are driven by differences in the underlying assumptions on emigration rates, mortality rates, and survey undercount of the foreign-born and undocumented immigrant population.

Agreement on this number is not universal. Rather than the survey based approach, Fazel-Zarandi et al. (2018) simulates year-over-year population changes by combining separate estimates of population inflows and outflows. Fazel-Zarandi et al. (2018) assume probability distributions around each inflow and outflow component and simulate the model over a range of values. The mean estimate based on the simulation model is 22.1 million undocumented immigrants for the year 2016 or twice the currently accepted estimate. The simulation produced a wide 95% confidence interval of as low as 16.2 million to as high as 29.5 million. This estimate has been highly criticized as being based on flawed assumptions on the rate of emigration by undocumented immigrants (Capps et al., 2018; Baker, 2021).

While I do not make any comment on the accuracy or inaccuracy of the assumptions of Fazel-Zarandi et al. (2018), this debate provides an example of *duelling certitudes* as defined by Manski (2011): Capps et al. (2018) and Baker (2021) question the assumptions asserted by Fazel-Zarandi et al. (2018) for assumptions that they prefer rather than on the methodology used. Recently, Van Hook et al. (2021) measured the uncertainty in the size of the unauthorized population from uncertainty in the underlying assumptions about coverage error, emigration, and mortality. Uncertainty in all three assumptions lead to a range in the size of the unauthorized foreign-born population between 9.1 and 12.2 million with 95% confidence. Van Hook et al. (2021) does not measure the uncertainty caused by item nonresponse as is done in this paper and which I find to a bigger source of uncertainty. In this analysis, I do not need to make any assumptions on the rate of emigration, rate of mortality, and rate of deportations.

Disagreement on the size of the undocumented population is more sever among politicians and the general public. In 2016, then presidential candidate Trump told supporters during a rally in Arizona "Honestly we've been hearing that number for years. It's always 11 million. Our government has no idea. It could be 3 million. It could be 30 million." The media response to this quote demonstrates the level of incredible certitude in current estimates of the size of the undocumented population with PolitiFact labeling this statement as "Pants on Fire" (The statement is not accurate and makes a ridiculous claim). Grigorieff et al. (2020) asked 1,193 people living in the US to estimate the proportion of immigrants and undocumented immigrants in the US.[4] Survey respondents overestimated the percentage of immigrants in the US by more than 20 percentage points from the Census estimates (34.7% vs 13.69%). Survey respondents also believed the undocumented population is 7.7 times larger than the estimates produced by the residual method at 25.4% of the total US population (Grigorieff et al., 2020). Whatever the true numbers are, there is a clear discrepancy in the perceived number of total foreign-born and undocumented immigrants in the US by the general public and the point estimates produced by government agencies and public policy institutions.

[4] This sample of 1,193 people living in the US was obtained as a non-probability quota sample to match the U.S. population in terms of age, gender, and region of residence

In this paper, I first document the trends and patterns in item nonresponse along with the imputation method used for the citizenship question in the ACS from 2009 to 2019. Item nonresponse has grown rapidly over the past 10 years with the largest increase in 2013. The sharp increase in 2013 coincides with the addition of the internet response mode and a reduction in the number of Failed-Edit Follow-Up calls. Individual respondents who did not respond to the mail/internet response modes and later respond to the phone or in-person interview are about 6 percentage points more likely to be foreign-born than those who responded through mail/internet. Phone or in-person respondents are about 6 percentage points more likely to be non-citizens than those who responded through mail/internet. This shows Foreign-born, in particular non-citizens who encompass all undocumented immigrants, are less likely to respond to the ACS and to the citizenship question.

I also show that item nonresponse is largest among those at or under the age of 18. While this age group has the largest nonresponse rates, citizenship status can be inferred through logical edits using their parents citizenship response. Roughly a quarter of non-respondents to the citizenship question can have their citizenship status logically edited using the response of their parents in a given sample year.

I estimate credible interval estimates of the size of the foreign-born population and across legal status. For each population group, the credible interval widens over time as item nonresponse in the citizenship question increases. The point estimates produced using the Census imputed values are near the lower bound estimate for the foreign-born and non-citizen credible interval estimates. This is what is expected when the MAR assumption is imposed and nonresponse in lower among native-born and naturalized foreign-born.

The estimated size of the foreign-born population could be as low as 12.3% or as high as 18.1% of the US population by 2019 compared to the 13.64% estimate produced when including imputed values.[5] The Census estimates that there are 328 million individuals residing in the US. With this total population estimate, the size of the foreign-born population in the US falls somewhere between 40.4 and 59.4 million compared to the Census estimate of 44.9 million. If the upper bound is the true population value, this would mean that as many as 14.6 million individuals are misclassifed as native-born citizens by the Cen-

[5] The Census value is slightly different than using the IPUMS files at 13.69%.

sus. At the lower bound, the Census could be overestimating the foreign-born population by 4.4 million.

In regards to the non-citizen population, the upper bound is 5.2 percentage points higher than the point estimates produced assuming nonresponse is MAR at 11.8% of the total US population in 2019. The lower bound of the non-citizen population is at 6.01% of the US population. This shares translate to a credible interval estimate of the size of the non-citizen population that fall between 19.7 and 38.7 million in 2019. This is in contrast to the estimated size by the Census of 21.7 million non-citizens residing in the US.

I next expand on the research estimating the size of the undocumented population by producing credible interval estimates of the undocumented population that take into account nonsampling error caused by item nonresponse in the citizenship question. I use the residual method proposed in Borjas (2017) and Borjas and Cassidy (2019) to identify undocumented immigrants in the ACS. The bounds of the credible interval estimate of the share of the undocumented population in the US are wide, ranging from a lower bound of 2.84% to an upper bound of 4.62% in 2019. This translates to a size of the undocumented population that falls between 9.3 and 15.2 million. In contrast, assuming nonresponse is MAR (i.e.using the Census' imputed values), the Borjas' residual method estimates the size of the undocumented population at 3.1% in 2019 or 10.15 million.[6]

The above credible interval estimates only take into account item nonresponse in the citizenship question. Multiple questions are used in the residual method to impute undocumented status at the individual level in the ACS with each having varying degrees of item nonresponse and different methods in the imputation procedure used to assign a status to item non-respondents. Item nonresponse from each questions exacerbates the issue of nonsampling error leading to wider interval estimates than if only focusing on nonsampling error caused by item nonresponse in the citizenship question. When taking into account item nonresponse from all questions used in the imputation procedure to assign legal status, the size of the undocumented population fall between 7.3 and 23.3 million.

The credible interval estimates can not exclude the possibility that the size of the undoc-

[6]My 2017 estimate is consistent to that of Passel and Cohn (2019) at 10.46 million compared to their 10.5 million estimate.

umented population has decreased, stayed flat, or increased over time. This is in contrast to Passel and Cohn (2019), Warren (2020), and Baker (2021) who have estimated a small decrease in the size of the undocumented population after 2008. The credible interval estimates show that, without making the critical and untested assumption that nonresponse is conditionally random, there is significant uncertainty in the estimated size of the undocumented population. Any value within this range is a credible estimate of the size of the undocumented population. Without making any assumptions about the exact distribution of the legal status of non-respondents, I cannot reject the lower end estimates produced by Fazel-Zarandi et al. (2018) that estimated the size of the undocumented population to be between 16.2 and 29.5 million.

These results have important implications for all work in the field of the economics of immigration that use survey data such as the ACS. Using the ACS estimates of the size of the foreign-born population or using a sample that includes non-respondents implicitly accepts the MAR assumption underpinning the imputed citizenship values. Error in these estimates and imputed values will bias the estimated effects whether studying the effect of immigration on an outcome of interest or the effect of immigration policy on immigrant outcomes.

This paper continues as follows. In Section 3.2, I detail the degree of item nonresponse to the citizenship question in the ACS and provide suggestive evidence that the conditional random assumption used by the Census to impute missing citizenship data may not be satisfied. In Section 3.3, I detail the residual method and the version used to identify undocumented immigrants in this paper. The Manski approach used to create the credible interval estimates are described in Section 3.4. The interval estimates of the size of the foreign-born population, the noncitizen population, and the undocumented population are presented in Section 3.5. Section 3.6 concludes and provides policy and empirical recommendations.

3.2 Item Nonresponse in the American Community Survey

The American Community Survey is the largest representative sample of the US, sampling 1% of households each year. It provides current demographic, social, economic, and housing information about the communities in America each year since its full implementation in 2005. The ACS is the predominate survey used by the Census to produce official yearly statistics of the total size of the foreign-born population and its legal composition (naturalized and non-citizen). As the ACS is the key data source used to estimate the count of the foreign-born population and used extensively as to produce estimates of the size of the undocumented population in the residual method (Passel and Cohn, 2019; Baker, 2021; Warren, 2020), I use the ACS for this analysis. The ACS data is sourced from IPUMS (Ruggles et al., 2020). This analysis focuses on the survey years 2009 to 2019. Survey year 2019 is the most recent year available. Survey year 2009 is the first year where a question on Medicaid participation is asked in the ACS, one of the questions used to identify legal immigrants in the residual method.

3.2.1 Census Sampling and Interview Process

The Census uses standard sampling methods to obtain its data.[7] The Census Bureau uses its Master Address File, which is composed of all known housing units and group quarters, to identify the household and group quarters that will be chosen for the sample that year. The ACS collects data each month of the year. The ACS yearly data files represent the average demographics as of July 1st of each year. Each month, the Census sends out requests for response. A household can respond through the paper questionnaire and, as of 2013, through the internet. The Mail and internet response modes are both known as self-response modes. Non-respondents to the self-response modes are then contacted for a computer-assisted telephone interview (CATI) the following month. In the third month, a third of non-respondents to the self-response and CATI modes are contacted in person to complete the ACS through a computer-assisted in-person interview (CAPI). In 2014, the Census reported that 65.5–68.7% of the addresses selected for the sample completed

[7] For more information on the data collection and ACS sample panels, see Bureau (2014).

the survey through the self-response modes and 96.7–98.0% of those contacted in person completed the survey (Bureau, 2014).

The American Community Survey (ACS) also conducts a follow-up operation to re-contact responding households to try to collect information missing or inconsistent in the mail and internet questionnaires to deal with survey nonresponse (Clark, 2014). This operation is called Failed Edit Follow-up (FEFU) calls. The FEFU calls are only for some households that self-responded to the survey and are conducted by phone interview (Clark, 2014).[8]

3.2.2 Citizenship Question

The Census uses the citizenship question to distinguish individuals as native-born or foreign-born. As well as to distinguish between naturalized citizens and non-citizens among the foreign-born. Figure 3.1 shows the citizenship question in the 2019 ACS. There are 5 options; (1) Yes, born in the United States, (2) Yes, born in Puerto Rico, Guam, the US Virgin Islands, or North Marianas, (3) Yes, born abroad of US citizen parent or parents, (4) Yes, US citizen by Naturalization, (5) No, not a US citizen. Choices (1), (2), or (3) are classified as native-born; while choices (4) or (5) are classified as foreign-born. The IPUMS files do not distinguish between category (1) and (2).

Figure 3.2 shows the share of the ACS sample that did not respond to the citizenship question. Item nonresponse for this question has increased from 2.07% of the sample in 2009 to 7.42% in 2019. A notable trend break appears in the year 2013. This is driven by changes in the survey collection methodology (addition of the internet response mode) and a reduction in the number of FEFU calls due to budgetary reasons (Clark, 2014).

Figure 3.3 shows the share of the population that is foreign-born by response status to the citizenship question.[9] The share of foreign-born among respondents has increased slightly from 2009 to 2019. Figure 3.4 show share of the population that are non-citizens by response status to citizenship question over time. The share of non-citizens among respon-

[8]FEFU are predominately done when mail respondents indicate that there are more than 5 individuals in the household as the mail questionnaire has only room for only 5 individuals (Clark, 2014).

[9]When discussing population share estimates, I use the weighted share. The weights are the person weights provided by the Census.

dents has decreased slightly over time, predominately after 2016. The share of foreign-born and non-citizens drops drastically among non-respondents in 2013 following the methodological changes to the ACS mentioned earlier. this is caused by the Census's 'hot-deck' imputation method that assumes the distribution of citizenship status is conditionally random. If non-response is not MAR, the imputation procedure will lead to a higher share on non-respondents to be improperly imputed as natives or naturalized citizens.

From the three figures above, the methodological changes to the ACS by the Census clearly had an impact on the degree of nonresponse to the citizenship question and on the share of non-respondents imputed as foreign-born and non-citizen. The methodological changes in 2013 did not have an impact on the response rate of other demographic questions such as race, Hispanic origin, sex, age, nor housing tenure questions (O'Hare, 2018; Clark, 2014). These questions are asked before the citizenship question so ordering might be an issue on item nonresponse when including sensitive questions on surveys.

I further disaggregate the above statistics by response mode. Figure 3.5 shows the share of the ACS sample by response mode.[10] The IPUMS files do not separate CATI and CAPI interviews. The share of the sample that responds by mail dropped by 40 percentage points after the introduction of the internet response model. As the share of individuals responding to the ACS through the internet mode increase, the share of the sample that responded by both mail and CATI/CAPI has decreased.

Figure 3.6 shows the share of item nonresponse to the citizenship question by response mode. The share of item nonresponse among mail respondents saw an almost doubling after the reduction in FEFU operations. Internet mode respondents had a slightly higher rate of item nonresponse. This is expected as they are less likely to be chosen for a FEFU. CATI/CAPI respondents did not see a trend break in 2013 but have seen a near tripling in item nonresponse to the citizenship question from 2009 to 2019. This may indicate that asking about citizenship status has become a more sensitive topic. As more individuals respond to the ACS through the Internet mode the issue of item nonresponse may worsen.

Figure 3.7 shows the share of the population that is foreign-born by response status to the citizenship question and response mode. Figure 3.8 shows the share of the population

[10] I exclude group quarter respondents.

that is non-citizen by response status to the citizenship question and response mode. Individual respondents who did not respond to the self-response modes and are later chosen to be interviewed by CATI/CAPI are about 6 percentage points more likely to be foreign-born than those who responded through the self-response modes. CATI/CAPI respondents are about 6 percentage points more likely to be non-citizens than those who responded through mail/internet. This shows Foreign-born, in particular non-citizens who encompass all undocumented immigrants, are less likely to respond to the ACS and to the citizenship question.

3.2.3 Imputation Procedure

The Census does not provide the exact methodology used for imputing citizenship status but has provided this information to IPUMS. For a detailed overview of the imputation procedure provided by the Census and released by IPUMS see Appendix C.[11] When a survey participant does not respond to the citizenship question the Census imputes a value in one of two ways: A logical editing procedure or a 'hot-deck' procedure. First, the census attempts to logically edit the non-respondents status using information from additional questions in the survey or through parental linkages if at least one parent is present. For instance, if the individual responds to being born in the US by responding they were born in the US in the place-of-birth question, they are logically edited as being native-born. If place-of-birth is also missing but a parent is present and is a native-born, the individual is logically edited as being native-born.

If a legal status cannot be logically edited, then the Census performs what they call a 'hot-deck' imputation procedure. The Census imputes a response to a non-respondent based on the citizenship response of a respondent with the same age, race, and ethnicity. The Census also takes into account geography in their 'hot-deck' imputation procedure, choosing a respondent of similarly observables that is also in the same area as the non-respondent (Bureau, 2014).[12]

[11] I cannot be certain that the imputation procedure as released by the IPUMS details the entire procedure used by the Census to impute citizenship status. For instance, the procedure detailed by IPUMS does not include at what geographic level the Census uses for imputation.

[12] How granular this area is is unknown to the public. It may be at the Census tract or at the state level.

The 'hot-deck' procedure is where the key assumption that nonresponse is MAR is imposed. As such, for my credible interval estimates, I will only focus on those that had their status allocated through the 'hot-deck' procedure. For this analysis, I logically edit non-respondents as native-born if there is a parent in the household that responded to the citizenship question as being native-born. I also logically edit native-born status if a person reports being born in the US when asked their place of birth. This procedure assigns a citizenship status to 23.8% of all imputed values.[13] Every other non-respondent is classified here as a 'hot-deck' imputed value. This simplified logical editing procedure will likely miss some of the individuals the Census logically edited. I do not logically edit citizenship status if the parent is not a native-born. Also, it is not clear if the Census uses a parents imputed citizenship status to logically edit the citizenship status of an individual of if they only use parental links when the parent respondent to the citizenship question.

Figure 3.9 shows the share of the sample with a 'Hot-deck' imputed citizenship response based on the definition in the paragraph above. The Census has been able to logically edit almost the same share of non-respondents but the trend in 'hot-deck' imputations has continued to rise. The 2013 trend break is also prominent in the 'hot-deck' imputed values. Logical edits are not enough to to deal with the rising trend of nonresponse nor the rise in nonresponse caused by the reduction in FEFUs calls.

As parental linkage is the key method of logically editing citizenship status, this will lead to differential rates of editing based on whether a parent is present or not in the household. Figure 3.10 shows that item nonresponse is largest among those under the age of 18. While this age group has the largest nonresponse rates, citizenship status can be inferred through logical edits using their parents citizenship response leading to this age group having the lowest level of nonresponse that needs to be imputed through the 'hot-deck' procedure. More work needs to be done to understand why parents are willing to respond about their own citizenship status but not when it comes to their children's citizenship status.

[13] A total of 17.26% of non-respondents are assigned nativity based on their mother's nativity while the other 6.54% of non-respondents are assigned nativity based on their father's nativity.

3.2.4 The Assumption that Nonresponse is Missing at Random

The assumption that non-respondents and respondents will have the same legal status distribution conditional on age-race-ethnicity is difficult to accept. For instance, this will require a native-born white Hispanic to have the same probability of item nonresponse as an undocumented white Hispanics of the same age. Due to the sensitivity of the legal status question, it is quite possible that the undocumented individuals will be less likely to respond than their native-born counterparts.

The MAR assumption can be seen most clearly in Figure 3.11 and 3.12. Figure 3.11 shows the distribution of foreign-born by response status across age for white Hispanics. Response status is now separated between those who responded or could have their citizenship status logically edited versus those that had their status imputed through the 'hot-deck' procedure. The distribution of foreign-born across age in non-respondents is similar but slightly higher than among respondents. The distributions do not perfectly match for two reasons. First, the Census imputation procedure takes into account geographic proximity. Item non-respondents are more likely to be in cities that have higher share of immigrants and non-citizens (O'Hare, 2018; Brown et al., 2018). Second, the differences in the logical editing procedure between my work and the Census official procedure. Figure 3.12 shows the distribution of being non-citizen across age for white Hispanics by response status. Both distributions are almost exact after age 15. The differences in the distribution below age 15 may be caused by the incomplete logically editing procedure relative to that used by the Census. The assumption that nonresponse is conditionally missing at random leads to an interesting outcome among the foreign-born population. In figure 3.13, the 'hot-deck' imputed foreign-born are less likely to be non-citizens than those who responded among white-Hispanics of all age groups.

3.3 Identifying Undocumented Immigrants

A major obstacle in estimating the size of the undocumented population is that large nationally-representative surveys do not ask respondents detailed questions on their documentation status. This has forced researchers and academics to create methods to infer documentation status in these surveys. The most popular of these methods is the residual method that was first developed by Warren and Passel (1987) to estimate the size of the undocumented population. The residual method estimates the size of the undocumented population by subtracting the estimated number of legal immigrants residing in the US from the estimated number of the total foreign-born population. The estimates for the total foreign-born population are derived from surveys that ask respondents where they are born such as the ACS. The number of legal immigrants is estimated using administrative data on legal admissions. This method has been further refined over the years to be able to assign undocumented status to individuals in large surveys such as the ACS based on demographic characteristics. Using the residual method, the estimated size of the undocumented populated was about 10.5 to 11.5 million people for the years 2017/2018 (Passel and Cohn, 2019; Baker, 2021; Warren, 2020). Estimates derived from the residual method have been widely used and are generally accepted as the best current estimates.

The Pew Research Center use a version of the residual method to identify undocumented immigrants in the ACS (Passel and Cohn, 2019). As this methodology underlies the "official" estimates reported by the DHS (Baker, 2021), I will focus on the detailing the methodology by the Pew Research Center only. There are slight differences in each of the different estimates of the size of the undocumented population. Differences in the method are primarily due to differences in assumptions of the undercount of the foreign-born population and assumptions about mortality and migration rates.

In rough terms, the methodology identifies the foreign-born persons in the sample who are likely to be legal using logical edits based on the individual's demographic, social, economic, and geographic characteristics and then classifies the remainder as likely to be undocumented. Passel and Cohn (2019) then apply a final filter to ensure that the counts from the micro-data agree with the official counts from the DHS of the total legal permanent resident (green card holders) and the official count from the DHS and the Department of State on legal non-permanent population (such as refugees, students with visas, workers with H1B visas, and tourist visas) through probabilistic methods that randomly assign legal or unauthorized status to those identified as potentially unauthorized individuals. Finally, the weights are adjusted to account for the estimated undercount of the undocumented population.

An issue for researchers is that the code underlying the undocumented status identifier is not publicly available. Borjas (2017) and Borjas and Cassidy (2019) “reverse engineered” the residual method by Pew Research Center to create a comparable identifier in all CPS and ACS files to identify likely undocumented immigrants. Borjas (2017) argues that only a few number of characteristics “matter” when it comes to identifying undocumented immigrants. A foreign-born individual is classified as a legal immigrant if any one of the following conditions are met;

1. that person arrived before 1980;[14]
2. that person is a citizen;
3. that person receives Social Security benefits, SSI, Medicaid, Medicare, or Military Insurance;[15]
4. that person is a veteran, or currently in the Armed Forces;
5. that person works in the government sector;
6. that person resides in public housing or receives rental subsidies, or that person is a spouse of someone who resides in public housing or receives rental subsidies;[16]
7. that person was born in Cuba;[17]
8. that person's occupation requires some form of licensing (such as physicians, registered nurses, air traffic controllers, and lawyers);[18]
9. that person's spouse is a legal immigrant or citizen.[19]

[14] Nearly all undocumented immigrants that arrived prior to 1980 are assumed to be legal as the majority were legalized through the IRCA 1986 reform and are assumed to have had enough time to change their legal status, migrate back to their home country, or died.

[15] Medicare and Medicaid information is only available for the years after 2007

[16] Information on public housing and rental subsidies is only available in the CPS and not the ACS.

[17] Practically all Cubans were granted refugee status through the Cuban Adjustment Act of 1966 and the wet feet, dry feet policy in 1995.

[18] Borjas (2017) does not detail exactly which occupations are used for this condition.

[19] For children living at home, this condition is expanded to include the parent's legal status as US laws allows under-aged children to have the same legal status as their legal parents.

The residual group of all other foreign-born persons is then classified as undocumented. This residual method provides comparable characteristics to those by the Pew Research Center and the "official" numbers produced by the DHS with the benefits that a researcher does not have to re-weight the data for the estimated undercount mentioned nor use probabilistic random assignment of status to match predetermined estimates. Due to its simplicity and detailed methodology which allow for replication, I identify undocumented immigrants in the ACS using Borjas and Cassidy (2019) residual method.

It is important to note that the validity of the assigned legal status is dependent on the validity of the procedure used to assign an individual's undocumented status in micro-data surveys. The accuracy of the assignment depends on the accuracy of the variables used to identify undocumented individuals. This paper does not focus on whether this method perfectly identifies undocumented immigrants given the logical editing procedure. This paper is only looking the magnitude of nonsampling error in the estimation of the size of the undocumented population caused by item nonresponse in the survey questions used to assign undocumented status under the assumption that the editing procedure of the residual method is accurate.

This is also just one, albeit the most popular, method to assign legal status. Another method uses the Survey of Income and Program Participation (SIPP) to assign legal status in the ACS using a two sample approach, usually referred to as the multiple imputation procedure. Ro and Van Hook (2021) compares the Borja's residual method with the multiple imputation procedure using the restricted version of the SIPP. Ro and Van Hook (2021) find notable demographic differences across both methods. While this method is not a focus of this paper, it will still be plagued by nonsampling error from item nonresponse. In fact it will be more severe as item nonresponse from both samples will need to be taken into account.

The residual method above produces an estimate of the size of the undocumented population of 3.1% or 10.15 million as of 2019. Figure 3.14 displays the share of the US population that is identified to be undocumented by their citizenship response status. As with the foreign-born and non-citizen populations, the share of undocumented immigrants among the item nonresponse group drops drastically in 2013 following the methodological changes to the ACS mentioned earlier. The share of undocumented immigrants for non-respondents is lower compared to the share of undocumented immigrants among only respondents for 2013 and beyond. As with the foreign-born and non-citizen shares in the non-respondent group, this is caused by the Census's imputation method assuming the distribution of citizenship status is MAR. This is counter to the intuition that undocumented immigrants are the least likely to respond sensitive questions on legal status and participate in surveys.

Figure 3.15 shows the share of the undocumented population by response mode and response status to the citizenship question. The share of the population identified as undocumented that responded to the citizenship question is very different across response modes. Citizenship question respondents are twice as likely to be identified as undocumented if they responded by internet compared to if they responded by mail (2% of internet respondents vs 1% of mail respondents). Among cati/capi respondents, about 5..5% of those that responded to the citizenship question are identified as being undocumented. In all, undocumented immigrants seem hesitant to participate in the ACS. Without phone or in-person follow up of self-response non respondents, the ACS would significantly underestimate the undocumented population.

3.4 Manski's Interval Estimation Method

Nearly all imputation models, including those used by the Census, assume nonresponse is MAR. In the case of the citizenship question in the ACS, conditional on a small set of observables (age,race, and ethnicity), the distribution of the foreign-born population among non-respondents is the same as respondents. This assumption allows point estimates to be produced. This assumption is a strong one. Without assuming the distribution of foreign-born status among non-respondents, only an interval estimate can be produced. Below I detail how the interval estimates are produced for the foreign-born, non-citizen, and undocumented population following Manski (2016).

For simplification, suppose that all population units are sample members. By the Law of Total Probability the share of the US population that are foreign-born can be defined as:

$$P(F) = P(F|R=1) \cdot P(R=1) + P(F|R=0) \cdot P(R=0) \tag{3.1}$$

where F=1 (or 0) signifies the population unit is foreign-born (or native-born). R = 1 (or 0) if a population unit did (or did not) report citizenship status. The empirical evidence identifies $P(R)$ and $P(F|R = 1)$. There is no empirical information on $P(F|R = 0)$. Without assuming the exact distribution of foreign-born status among non-respondents $P(F|R = 0)$ can take any value between 0 and 1. This yields the following sharp bounds:

$$P(F|R=1) \cdot P(R=1) < P(F) < P(F|R=1) \cdot P(R=1) + P(R=0) \tag{3.2}$$

To estimate the lower bound, one supposes that F = 0 for each sample member with missing data in the citizenship question. To estimate the upper bound, one likewise supposes that F = 1 whenever observation is missing. Thus, estimation of the bound simply requires two extreme imputations of each case of missing data. The point estimate of the share of the foreign-born when using the Census imputations lies between the upper and lower bounds.

The same procedure can be used to create interval estimates for the share of non-citizens. The interval estimates can be written as:

$$P(NC|R=1) \cdot P(R=1) < P(NC) < P(NC|R=1) \cdot P(R=1) + P(R=0) \tag{3.3}$$

where NC = 1 (or 0) if the population unit is a non-citizen (or not a non-citizen). R = 1 (or 0) if a population unit did (or did not) report citizenship status. At the upper bound of the estimates, all individuals who did not respond to the citizenship question are assigned as non-citizens. This also corresponds to where all non-respondents are foreign-born. In the lower bound, all non-respondents are allocated as not being non-citizens. This may correspond to either all non-respondents are native-born or naturalized foreign-born citizens. I allocate the non-residents as native-born so that the lower bound of the non-citizen interval estimates also corresponds to the lower bound of the foreign-born estimates.

There are additional steps to produce the interval estimates of the size of the undocumented population. The residual method uses multiple questions to assign legal status to each individual. The method also assigns legal status based on the legal status of an individual's spouse or parent if they are present in the household.

For simplicity, I first focus on item nonresponse of the citizenship question. To create a lower bound I assign all citizenship question non-respondents as native-born and run the residual method procedure. To create an upper bound I assign all non-respondents as non-citizens and run the residual method procedure.[20]

This is better than assigning all non-respondents as undocumented or not undocumented if they did not respond to the citizenship question. For example a non-respondent may have responded that they are currently in the armed forces. Assigning undocumented status to that individual would be improper. The above procedure ensures that the individual is logically edited as being documented regardless if assigned as a non-citizen.

We then produce interval estimates of the undocumented population taking into account nonsampling error caused by nonresponse to all questions used in the assigning procedure. At the lower bound I assign non-respondents an answer to all imputed questions that would logically edit the individual as being documented through the residual method. At the upper bound I assign non-respondents an answer to all imputed questions that would fail to logically edit the individual as a documented immigrant.

3.4.1 Assumptions

The key benefit of producing interval estimates is that I do not need to make assumptions about the non-response process and I are able to estimate the maximum nonsampling error caused by item nonresponse. Even so, as I am only focusing on nonsampling error from item nonresponse only, I must make key assumptions.

First, I do not take into account unit nonresponse. Instead, I assume the census weights accurately deal with unit nonresponse and use the census weights provided to create the interval estimates. Since unit and item nonresponse is greatest in areas with the highest share of non-citizens (Brown et al., 2018), there is likely nonsampling error in the weighting procedure as well. Taking into account unit nonresponse will lead to larger bounds than those estimated here.

[20] An alternative method to create interval estimates is to use the basic form of the residual method. That is, to subtract the lower and upper bound estimates of the foreign-born population from the estimated legal foreign-born population produced using administrative data. This would require assumptions to be made on the quality of the estimates from the administrative data. As the goal is to minimize the number of untestable assumptions made, I do not perform such exercise.

Second, I accept respondents answers as accurate. It has been documented that recent immigrants misreport their citizenship status (Van Hook and Bachmeier, 2013). Brown et al. (2018) shows non-citizens are significantly more likely to falsely report being citizens (both naturalized and native-born). In the 2016 ACS, 34.7% of respondents that are identified in administrative records as being non-citizens claim to be citizens. Brown et al. (2018) also show misreporting among all immigrants regardless of years in the US unlike Van Hook and Bachmeier (2013) that looked at differences in aggregate estimates. Of those that misreport their citizenship status, approximately 15.9% report being citizens from birth. Among respondents in the ACS linked to administrative records using Individual Taxpayer Identification Numbers (most likely to be undocumented), 11% said they were US citizens and 6.6% said they were native-born (Brown et al., 2018). There was virtually no misreporting among ACS respondents that had been identified as citizens in administrative records.

The logical editing procedure used by the Census magnifies misreporting error within households if one parents falsely reports being a citizen as all non-responding children in the household would be assigned citizenship status. This is also the case for the editing procedure in the residual method. This can cause major issues in the interval estimates produced here as I only study item nonresponse for those who's citizenship status cannot be logically edited - those assigned a citizenship status only through the assumption that nonresponse is conditionally random. The interval estimates produced here will therefore most likely be shifted down than if all individuals responded truthfully to the citizenship question. Given the evidence provided in Brown et al. (2018), the true size of the foreign-born, non-citizen, and undocumented populations would likely be closer to the upper bound and possibly exceed the upper bounds estimated here depending on the degree of error in the weighting procedure and degree of misreporting.

3.5 Credible Interval Estimates

3.5.1 Foreign-Born Estimates Recognizing Item Nonresponse

As discussed in Section 3.4, I create an upper bound estimate of the size of the foreign-born population by assigning foreign-born status to any individual that did not respond to the citizenship question and could not have their citizenship status logically edited. I create the lower bound estimate by assigning native-born status to any individual that did not respond to the citizenship question and could not have their citizenship status logically edited. Figure 3.16 shows the credible interval estimates of the share of the US population that are foreign-born. The long-dash line represents the upper bound of the share of foreign-born in the population where all 'hot-deck' imputed values are assigned as foreign-born. The short-dash line represents the lower bound of the share of foreign-born in the population where all MAR imputed values are assigned as native-born. The solid line represents the share of foreign-born in the population using the imputed values from the Census under the assumption that nonresponse is MAR.

The estimated size of the foreign-born population by 2019 could be as low as 12.3% or as high as 18.1% of the US population compared to the 13.64% estimate produced with Census imputed values. This is a significant degree of uncertainty and any value within these bounds can not be rejected without further making assumptions of the distribution of foreign-born among non-respondents. With an estimated 328 million individuals residing in the country, the upper bound would indicate a size of the foreign-born population at 58 million. If the upper bound is the true population value, this would mean that as many as 14.6 million individuals are misclassifed as native-born citizens by the Census. At the other lower bound, there may be only 40.3 million foreign-born individuals in the country or 4.4 million less than Census estimates. The Census estimates produced under the assumption of MAR are closer to the lower bound than the upper bound. That is due to the Census imputing a distribution of citizenship status among non-respondents that is conditionally the same as for respondents.

3.5.2 Non-citizen Population Estimates Recognizing Item Nonresponse

Figure 3.17 shows the credible interval estimates of the share of the US population that are non-citizens. The long-dash line represents the upper bound of the share of non-citizens in the population where all MAR imputed values are assigned as foreign-born. The short-dash line represents the lower bound of the share of non-citizens in the population where all MAR imputed values are assigned as native-born. The solid line represents the share of non-citizens in the population using the imputed values from the Census under the assumption that nonresponse is conditionally random

In regards to the non-citizen population, the upper bound is 11.8% of the total US population in 2019. The lower bound of the non-citizen population is at 6% of the US population. The Census estimates 6.6% of the population are non-citizens. This shares translate to a credible interval estimate of the size of the non-citizen population between 19.7 and 38.7 million in 2019. This is in contrast to the estimated size by the Census of 21.7 million non-citizens residing in the US as of 2019. Again, the Census point estimates are closer to the lower bound than to the upper bound for the same reason outlined above.

3.5.3 Undocumented Population Estimates Recognizing Item Nonresponse

Figure 3.18 shows the credible interval estimates from this procedure. The short dash lines represent the lower bound while the long dash lines represent the upper bound estimates. The bounds of the credible interval estimate of the share of the undocumented population in the US are wide ranging from a lower bound of 2.84% to an upper bound of 4.62% in 2019. This translates to a size of the undocumented population that falls between 9.3 and 15.2 million. Assuming nonresponse is MAR (solid line in Figure 3.18), the Borjas' residual method estimates the size of the undocumented population at 3.1% in 2019 or 10.15 million.[21] The credible interval estimates can not exclude the possibility that the size of the undocumented population has stayed flat, increased, or decreased over time.

The above credible interval estimates only take into account item nonresponse in the citizenship question. Multiple questions are used in the residual method to impute undoc-

[21] Our 2017 estimate is consistent to that of Passel and Cohn (2019) at 10.46 million compared to their 10.5 million estimate.

umented status at the individual level in the ACS with each having varying degrees of item nonresponse and different methods in the imputation procedure used to assign a status to item non-respondents. Item nonresponse from each questions exacerbate the issue of nonsampling error leading to wider interval estimates than if only focusing on nonsampling error caused by item nonresponse in the citizenship question.

Figures B.1 to B.7 look at the degree of nonresponse to all other question used in the imputation procedure.[22] Panel (A) of each figure looks at the share of the sample that do not respond to each question. I specifically look at the share of nonresponse on a sample composed of individuals that responded in the ACS that they are non-citizens and individuals that did not respond to the citizenship question. These are the individuals that compose the bounds for the size of undocumented population. As with the citizenship question, nonresponse to all other questions has grown rapidly over time. A similar spike in nonresponse appears after 2013 coinciding with the methodological changes to the ACS. This means less information is available to deduce documentation status from other questions. More specifically, to assign documentation status the residual method is using the imputed values from the Census. If there is error in the imputation procedure used by the Census in any of these questions, the residual method will be biased when assigning documentation status.

Panel (B) of Figures B.1 to B.7 show the share of the sample that satisfy the condition of interest by the response status to the Question of interest. For all except those born in Cuba, the sample with imputed individuals are more likely to be imputed a value that satisfy conditions that would assign them as legal immigrants in the Borjas residual method.

Figure B.8 shows the share of the sample that did not respond to at least one of the questions used in the residual method. Again, this is a sample of those that responded as non-citizens or that did not respond to the citizenship questions. Among this sample, around 65% of the sample did not respond to at least one of the questions by 2019. Roughly 50% of those that responded to all the questions used in the residual method are classified as being undocumented immigrants (Panel B). Among the sample that had at least one question imputed, the share undocumented was considerably lower at a little over 10%.

[22]Due to an error in the constructed imputation flag for Medicaid and Medicare produced by IPUMS, I exclude those questions from my analysis.

Next, I produce credible interval of estimates of the undocumented population recognizing nonresponse in all variables as described in Section 3.4. When taking into account item nonresponse from all questions used in the imputation procedure to assign legal status, the estimated size of the undocumented population fall between 7.3 and 18.3 million. This is a considerable expansion in the size of the bounds. While the bounds are wide, they are of great value as they include all possible uncertainty caused by item nonresponse. Future work is needed to reduce the size of the bounds.

3.6 Concluding Remarks and Recommendations

Estimates of the size and legal composition of the foreign-born population have been treated with incredible certitude by the media and policy analysis alike (Manski, 2016). These estimates contain both sampling and nonsampling error that are occasionally discussed but rarely ever estimated. This paper considers nonsampling error in these estimates caused by item nonresponse. Item nonresponse in the citizenship question used to derive these estimates has grown rapidly over the past 10 years. This is a considerable issue when estimating the size and legal composition of the foreign-born population as the imputation and weighting procedures used to deal with nonresponse assumes that nonresponse is conditionally random. An assumption that has empirically been put into question recently (Brown et al., 2018).

In this paper, I produced credible interval estimates of the size of the foreign-born, noncitizen, and the undocumented immigrant populations that takes into account nonsampling error from item nonresponse following the approach in Manski (2016). This approach produces the maximum degree of uncertainty in the estimates of these population groups caused from item nonresponse. Interval estimates of the size of the foreign-born population in the US is fall between 40.4 and 59.4 million as of 2019 compared to the Census estimate of 44.9 million. Interval estimates of the size of the undocumented population ranges between 9.3 and 15.2 million when taking into account nonresponse to the citizenship question only. This is compared to the widely accepted estimate of 11 million undocumented immigrants. When taking into account item nonresponse from all questions used in the imputation procedure to assign legal status, the size of the undocumented population fall between 7.3 and 18.3 million.

Taking into account item nonresponse creates considerable uncertainty in the estimates of the size and legal composition of the foreign-born population. The critical take from this exercise is that more research needs to be done to be able to produce accurate estimates the size of these hard to reach populations. This work highlights the significance of initial assumptions to reduce uncertainty from non-classical measurement error. It is important for those estimating the size of these population groups to justify the initial assumptions made used to derive their point estimates.

While the interval estimates are large, they provide a valuable baseline to alternative methods and assumptions used to estimate of the size and legal composition of the foreign-born population. Any estimate within these bounds should be considered credible estimates. That is, until the bounds are able to be reduced using additional information to supplement the survey data. While currently it may not be known the exact size of the undocumented population, this work gives some indication of what the right answer might be, and that there very well be far more undocumented immigrants in the US than currently estimated.

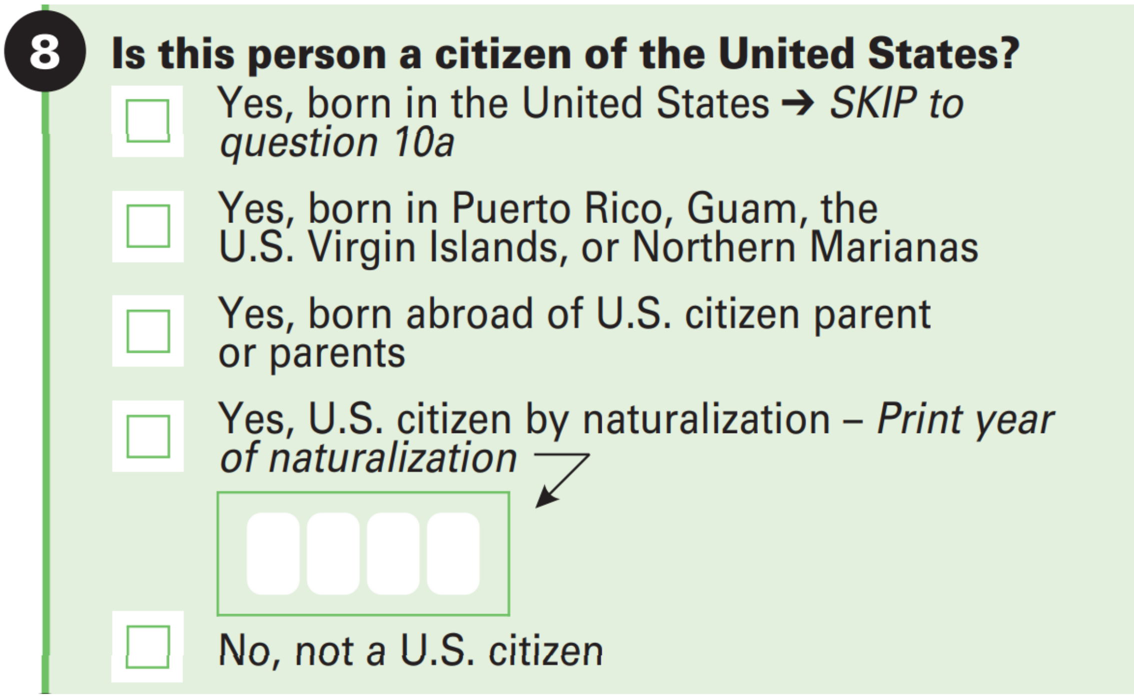
This is also of valuable interest to policy makers who use these estimates to predict the cost and benefits of immigration policy. Without making critical, and untested, assumption on the survey data or without having supplementary data, there will be considerable uncertainty on the estimated cost and benefits of any immigration policy.

Recommendations for the Census Bureau: The Census' decision to reduce FEFU calls is a significant reason for the large increase in item nonresponse documented for the citizenship question. To reduce the rate of nonresponse, the Census will need to increase the rate of FEFU calls. Putting aside the methodological changes, nonresponse to the citizenship question has continued to rise, indicating asking this question has become more sensitive over time. Increasing FEFU calls will still not deal with the continued rise in nonresponse nor in dealing with the problems of misreporting. The Census should consider improving the hot-deck imputation method by adding more covariates or modify the MAR assumption. Alternatively, the Census can link administrative records that contain legal status to assign citizenship status to non-respondents. This will considerably reduce the bounds estimated here. Additionally, linking to administrative records will help correct for the significant rate of misreporting documented in Brown et al. (2018) which was not considered in this paper.

Recommendations for Researchers: All work that uses Census data implicitly accepts the MAR assumption that the non-respondents citizenship status has the same distribution as respondents conditional on a set of observables. Using the Census estimates to estimate the effect of the size of foreign-born population on an outcome of interest may bias estimates. Using a sample that includes non-respondents to estimate the effect of immigration policy on immigrants may also lead to biased estimates. Given that the distribution of legal status across non-respondents is unknown, these results suggest work using the citizenship question should use bounding estimates. Aside from using the Census estimates, researchers should produce estimates using the lower and upper bound by assigning the extreme cases to the imputed values.

3.7 Chapter 3 Figures

Figure 3.1: ACS Citizenship Question

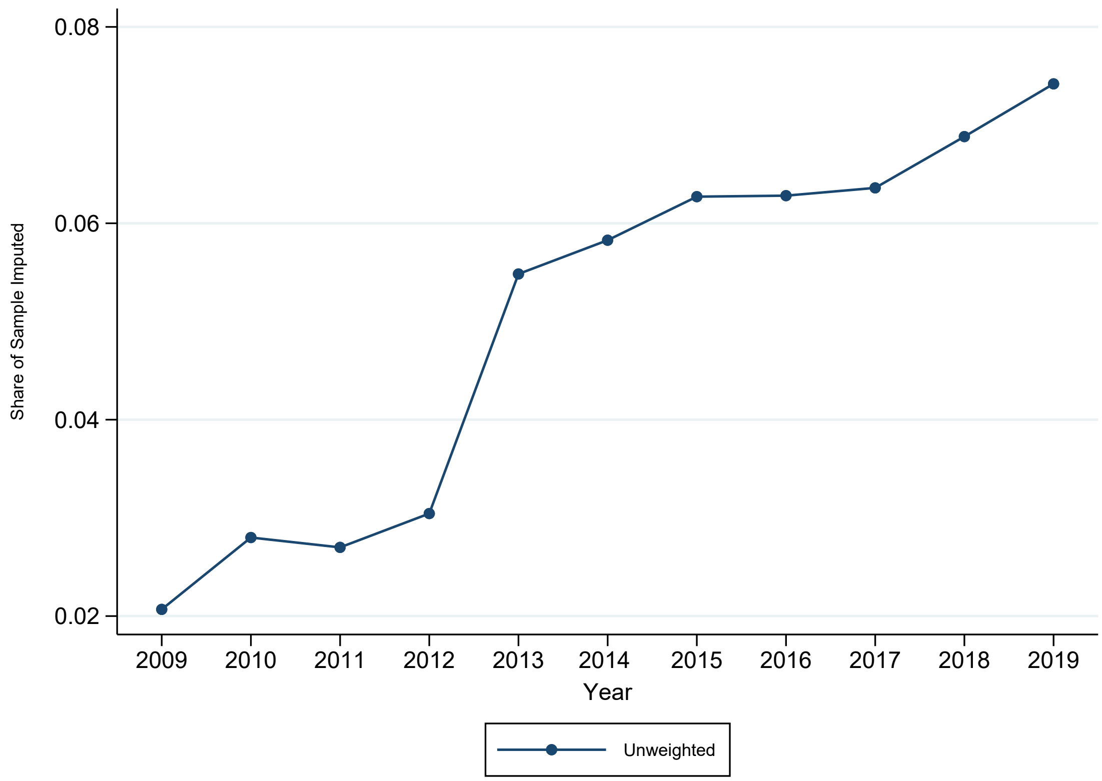


8 **Is this person a citizen of the United States?**

- Yes, born in the United States → *SKIP to question 10a*
- Yes, born in Puerto Rico, Guam, the U.S. Virgin Islands, or Northern Marianas
- Yes, born abroad of U.S. citizen parent or parents
- Yes, U.S. citizen by naturalization – *Print year of naturalization*
- No, not a U.S. citizen

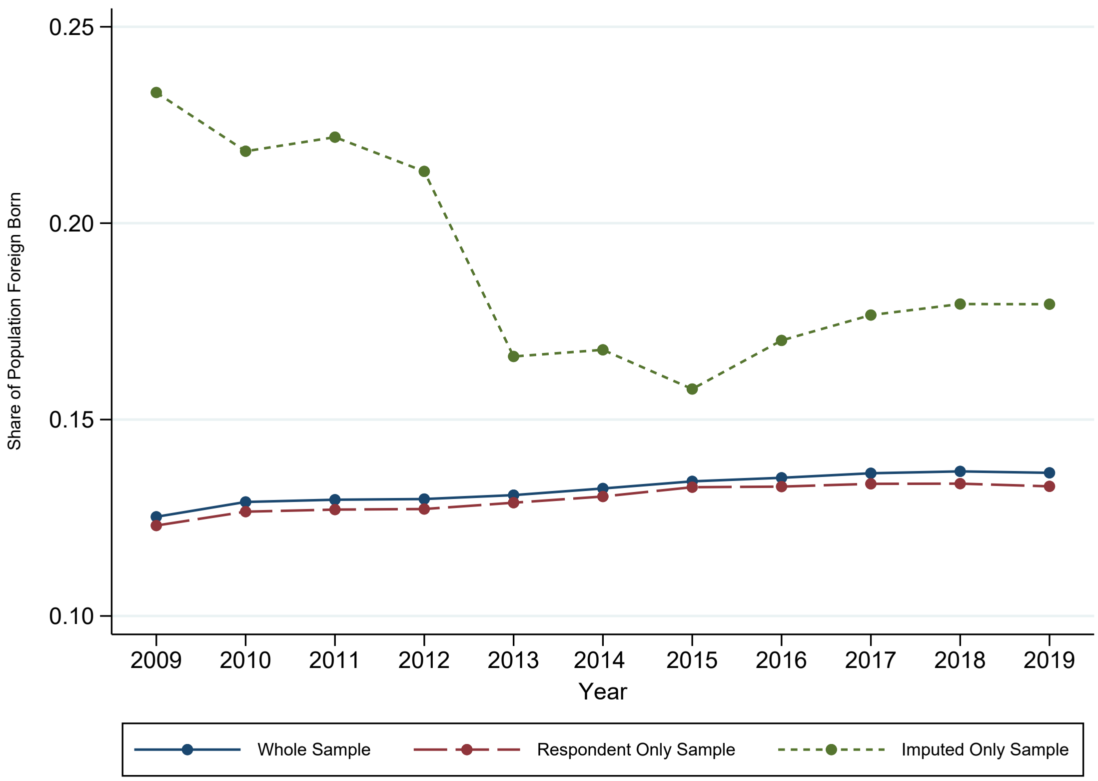
Notes: - 2019 American Community Survey Questionnaire

Figure 3.2: Share of Sample with Imputed Citizenship Question



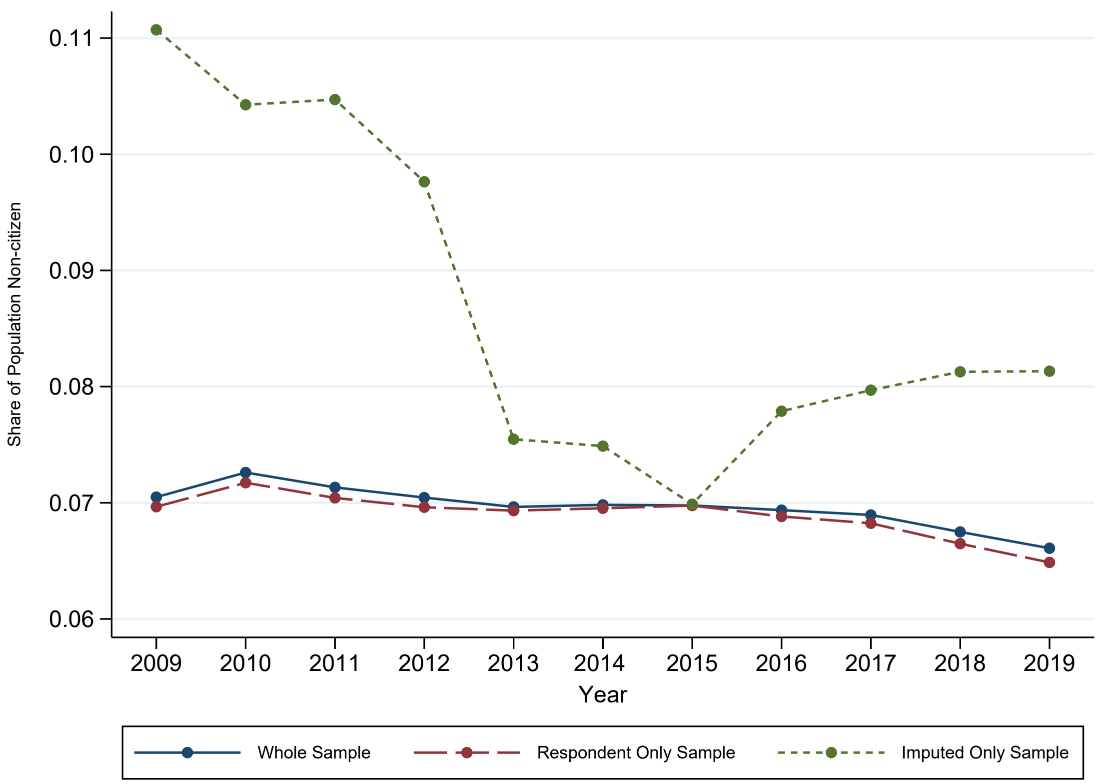
Notes: Authors own calculations using the American Community Survey. Shares are the unweighted raw totals.

Figure 3.3: Share of Population Foreign-Born by Response Status



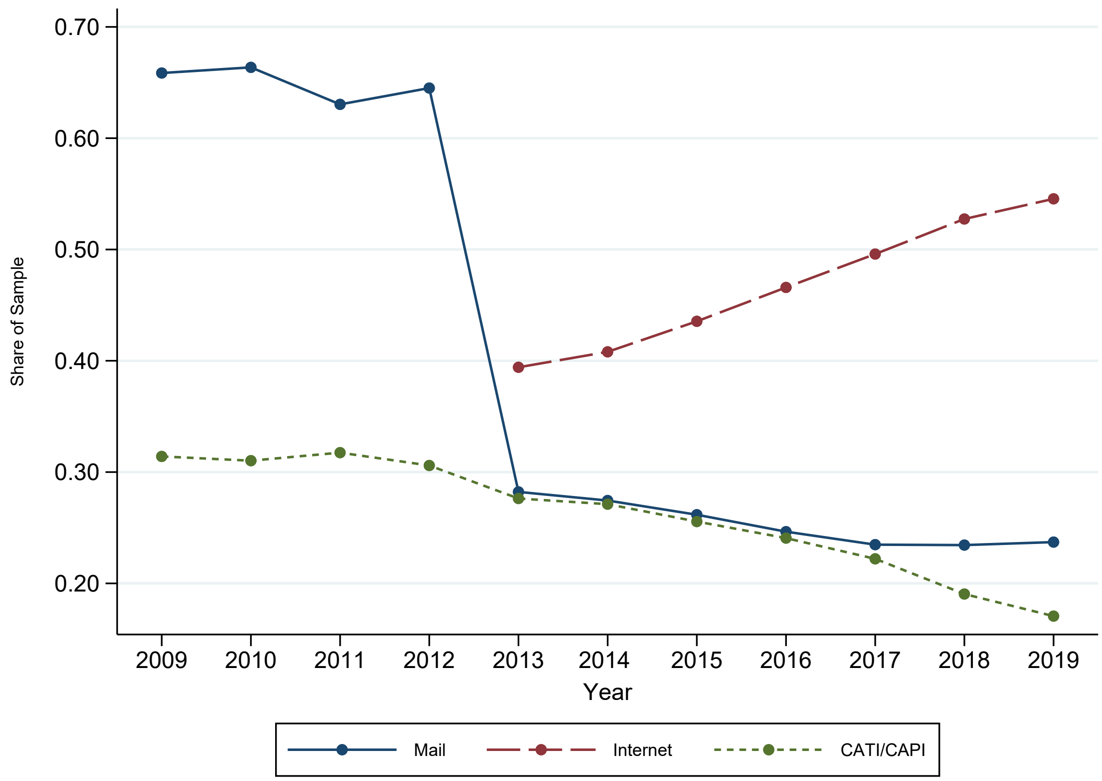
Notes: Authors own calculations using the American Community Survey. Weights used are person weight provided by Census.

Figure 3.4: Share of Population Non-Citizen by Response Status



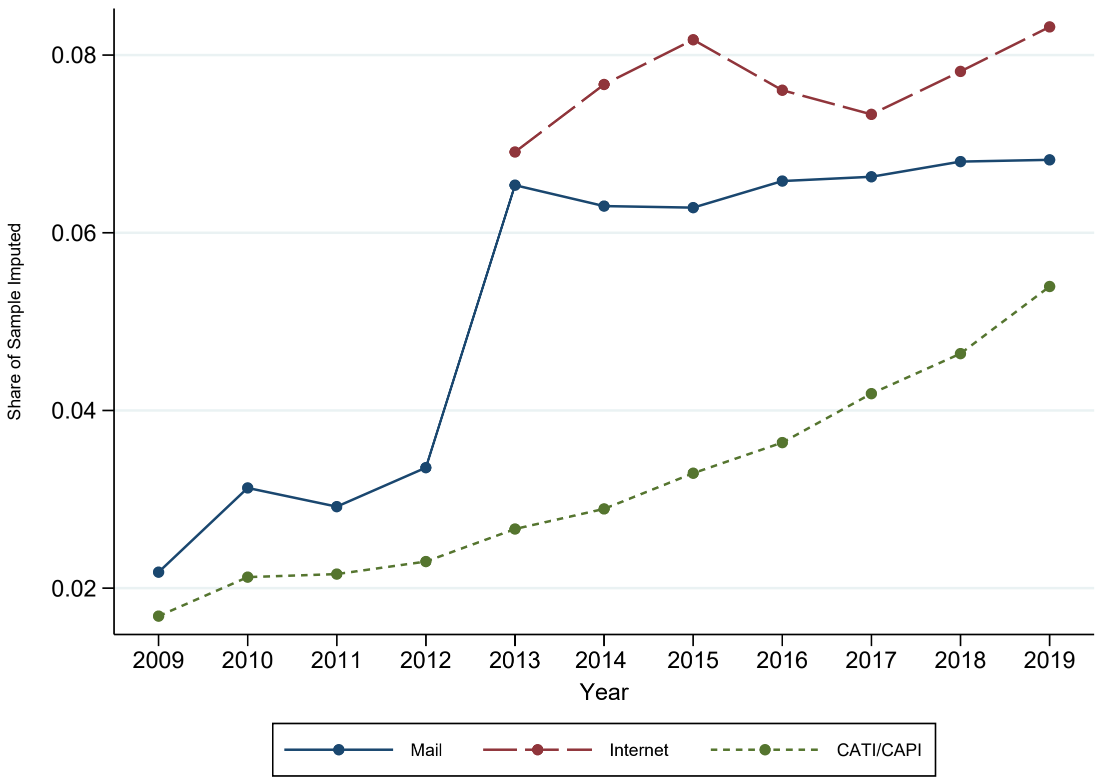
Notes: Authors own calculations using the American Community Survey. Weights used are person weight provided by Census.

Figure 3.5: Share of Sample by Response Mode



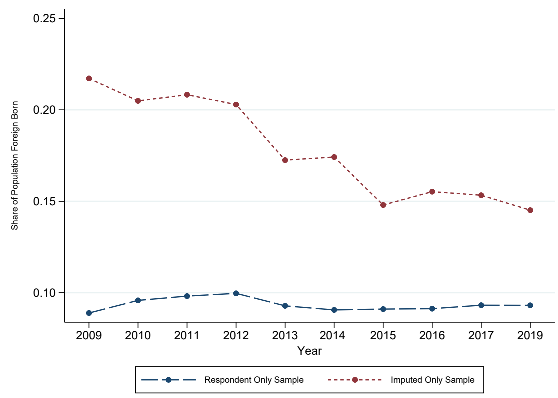
Notes: Authors own calculations using the American Community Survey. Shares are the unweighted raw totals.

Figure 3.6: Share of Sample with Imputed Citizenship Question by Response Mode

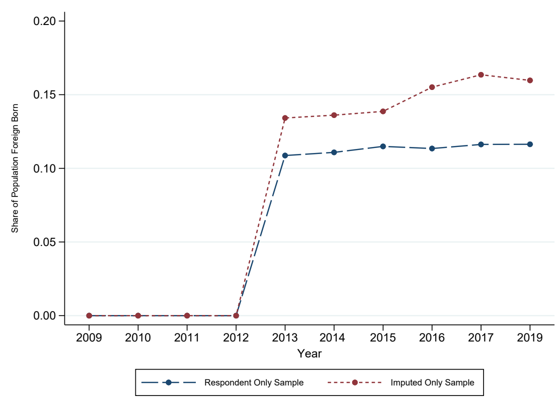


Notes: Authors own calculations using the American Community Survey. Shares are the unweighted raw totals.

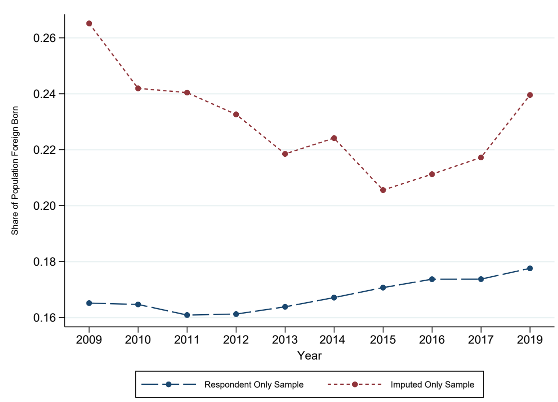
Figure 3.7: Share of Population Foreign-Born by Response Mode



(A) Responded by Mail



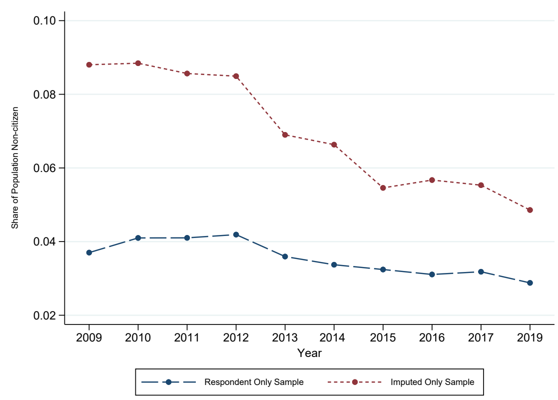
(B) Responded by Internet



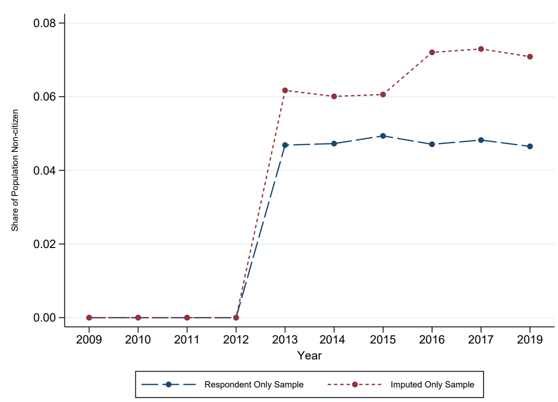
(C) Responded by CAPI/CATI

Notes: Author's own calculations using the American Community Survey. Shares are calculated using census person weights.

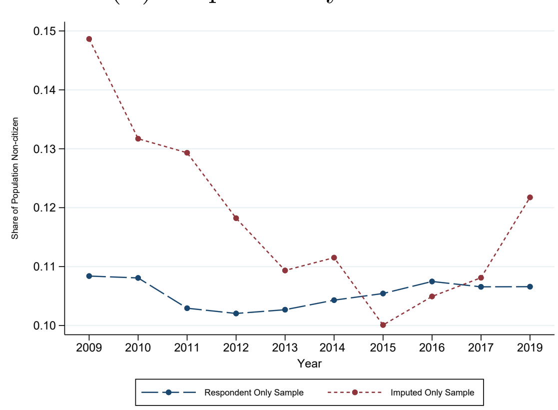
Figure 3.8: Share of Population Non-citizen by Response Mode



(A) Responded by Mail



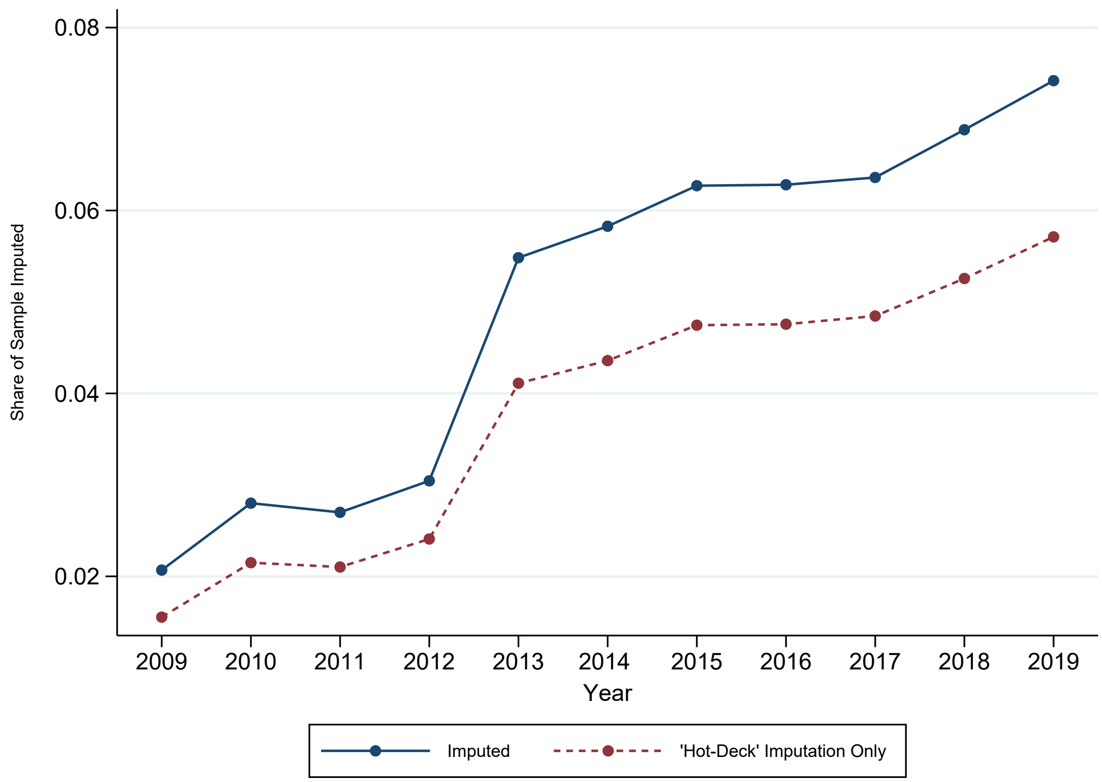
(B) Responded by Internet



(C) Responded by CAPI/CATI

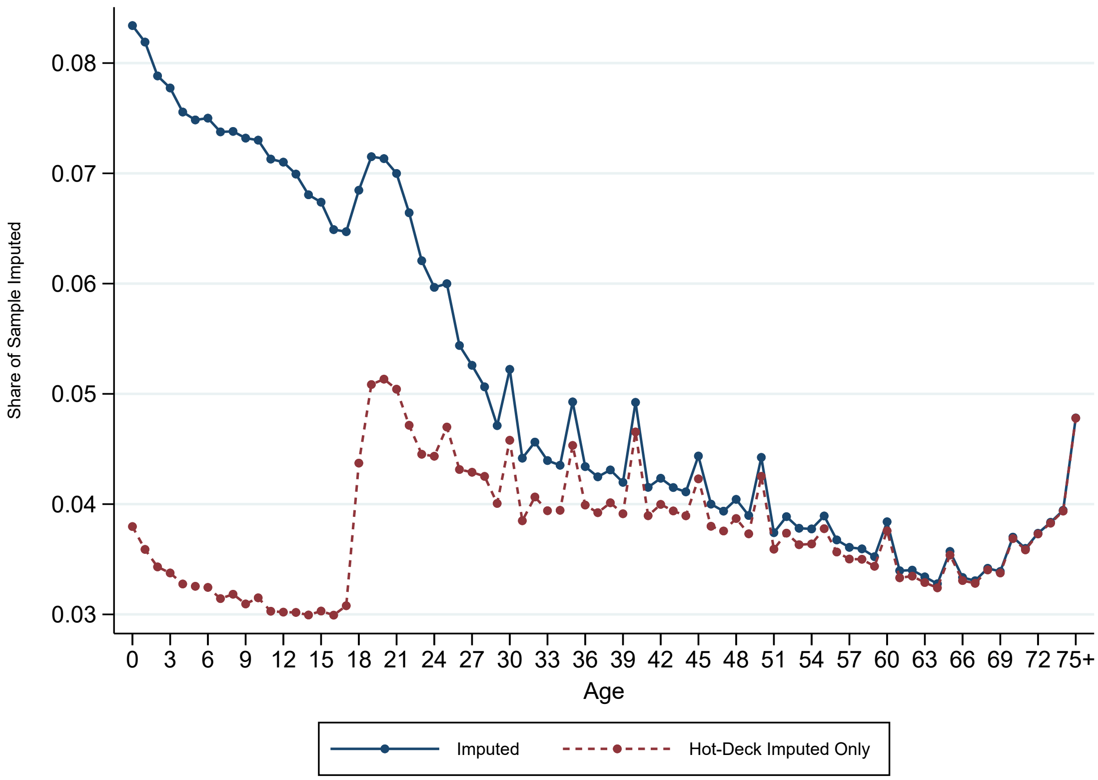
Notes: Author's own calculations using the American Community Survey. Shares are calculated using census person weights.

Figure 3.9: Share of Sample with 'Hot-Deck' Imputed Citizenship Question



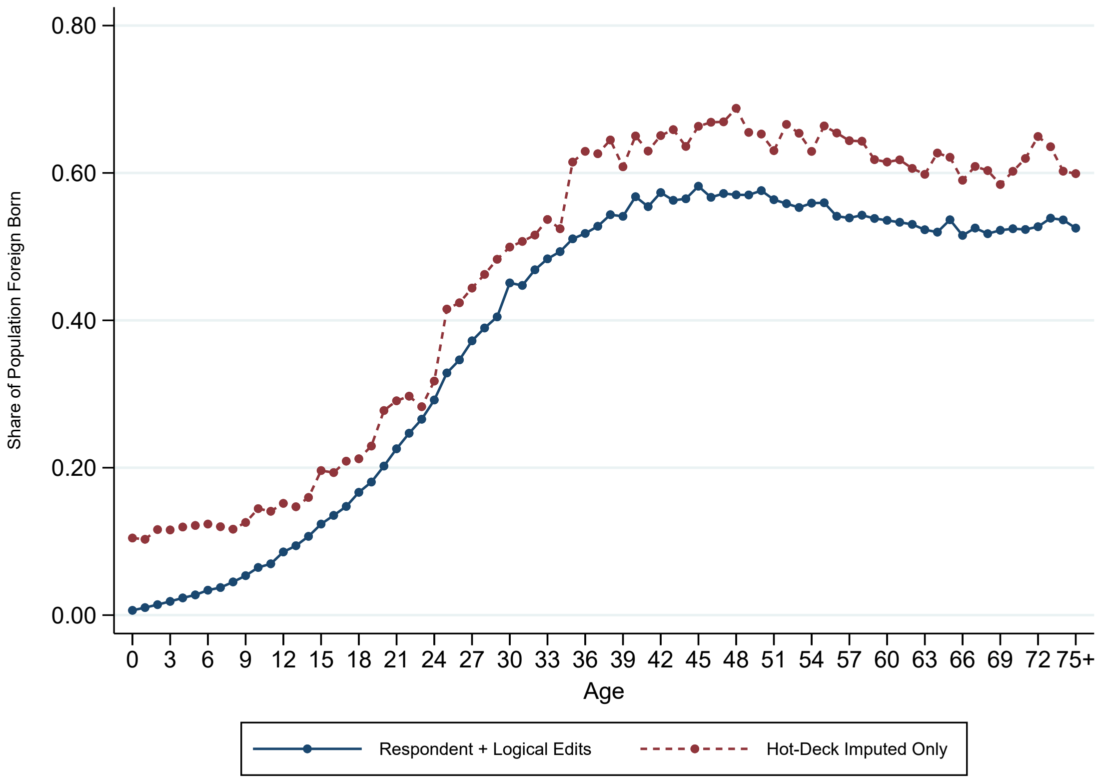
Notes: Authors own calculations using the American Community Survey. The solid blue line is the share of the sample that are flagged as not responding to the citizenship question. The dashed red line is the share of the sample where citizenship could not be logically edited from additional data in the survey. Both shares are the unweighted raw totals.

Figure 3.10: Item Nonresponse in Citizenship Question Across Age



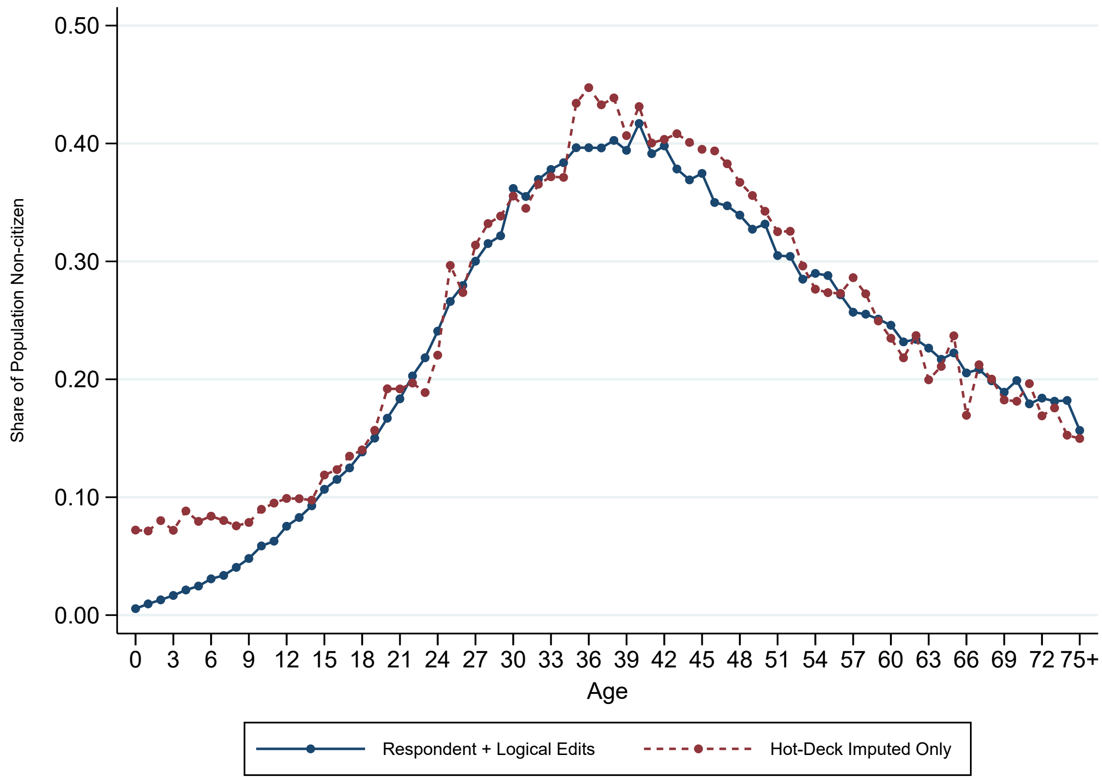
Notes: Authors own calculations using the American Community Survey. Figure combines the sample years 2009 to 2019. The solid blue line is the share of the sample that are flagged as not responding to the citizenship question. The dashed red line is the share of the sample where citizenship could not be logically edited from additional data in the survey. Both shares are the unweighted raw totals.

Figure 3.11: Share Foreign-Born by Response Status Across Age



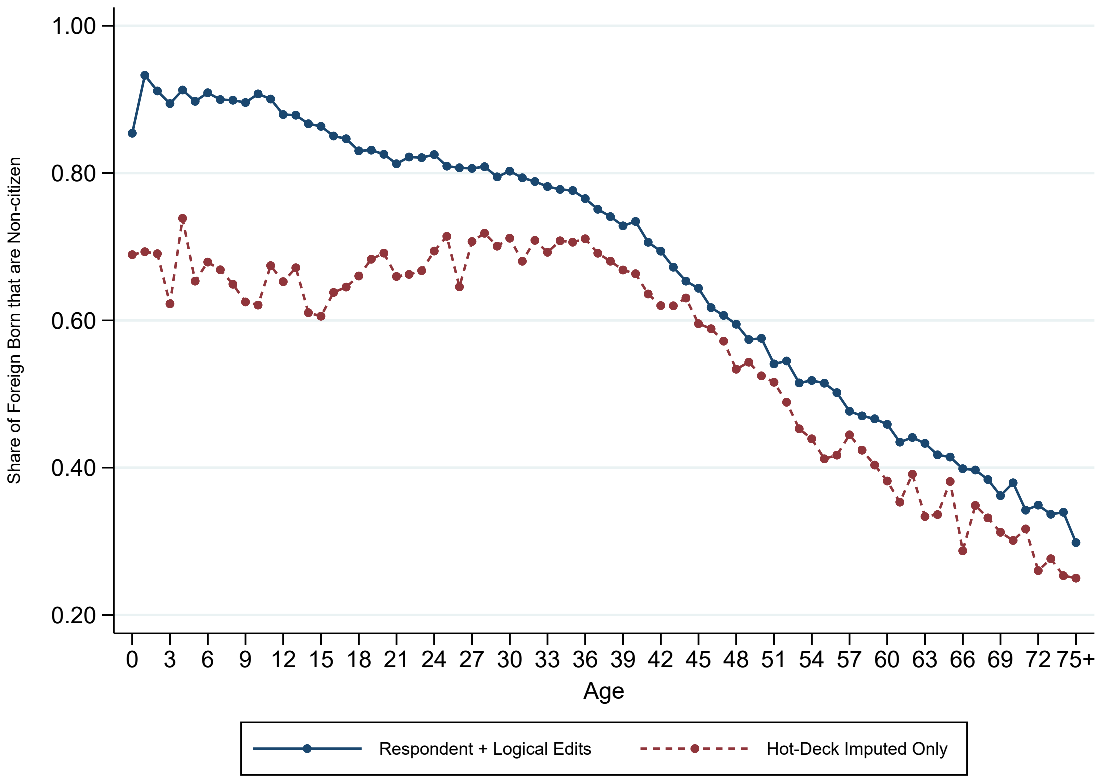
Notes: Authors own calculations using the American Community Survey. Figure combines the sample years 2009 to 2019. Sample restricted to white Hispanics. The solid blue line is the share of the sample that are flagged as not responding to the citizenship question. The dashed red line is the share of the sample where citizenship could not be logically edited from additional data in the survey. Both shares are the unweighted raw totals.

Figure 3.12: Share Non-Citizen by Response Status Across Age



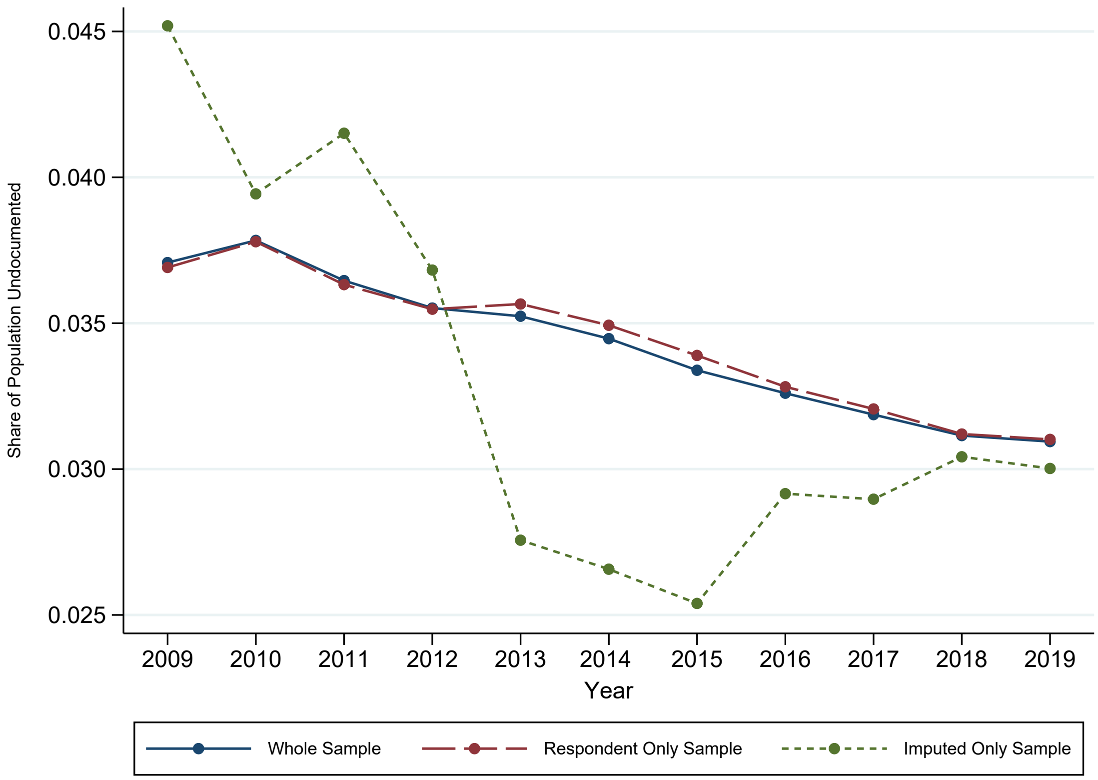
Notes: Authors own calculations using the American Community Survey. Figure combines the sample years 2009 to 2019. Sample restricted to white Hispanics. The solid blue line is the share of the sample that are flagged as not responding to the citizenship question. The dashed red line is the share of the sample where citizenship could not be logically edited from additional data in the survey. Both shares are the unweighted raw totals.

Figure 3.13: Share Foreign-Born That is Non-Citizen by Response Status Across Age



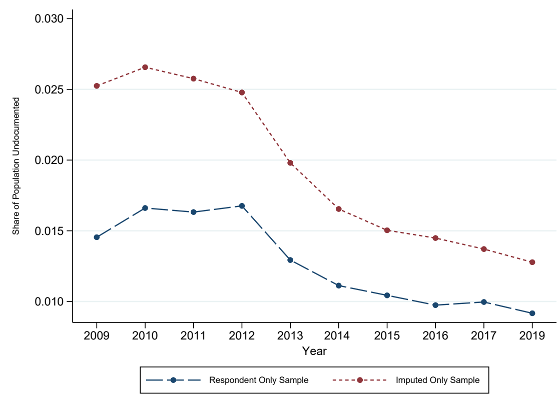
Notes: Authors own calculations using the American Community Survey. Figure combines the sample years 2009 to 2019. Sample restricted to white Hispanics. The solid blue line is the share of the sample that are flagged as not responding to the citizenship question. The dashed red line is the share of the sample where citizenship could not be logically edited from additional data in the survey. Both shares are the unweighted raw totals.

Figure 3.14: Share of Population Undocumented by Response Status

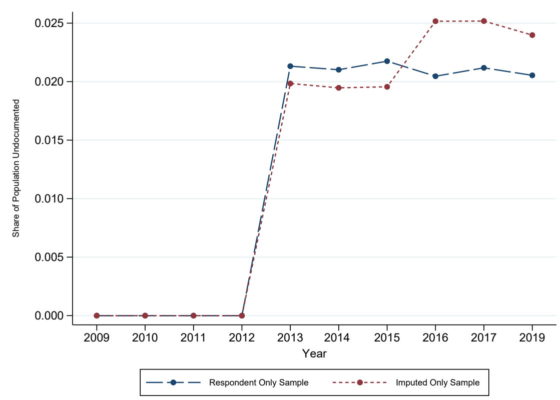


Notes: Authors own calculations using the American Community Survey. Weights used are person weight provided by Census.

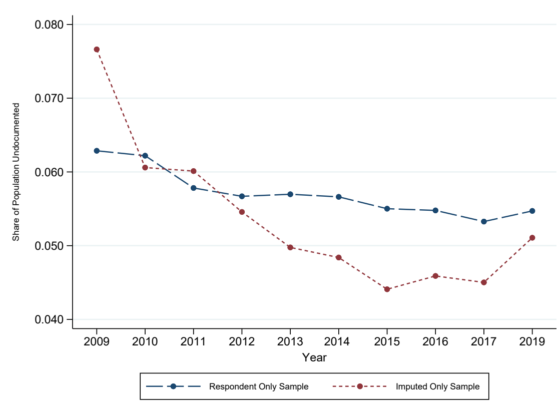
Figure 3.15: Share of Population Undocumented by Response Mode



(A) Responded by Mail



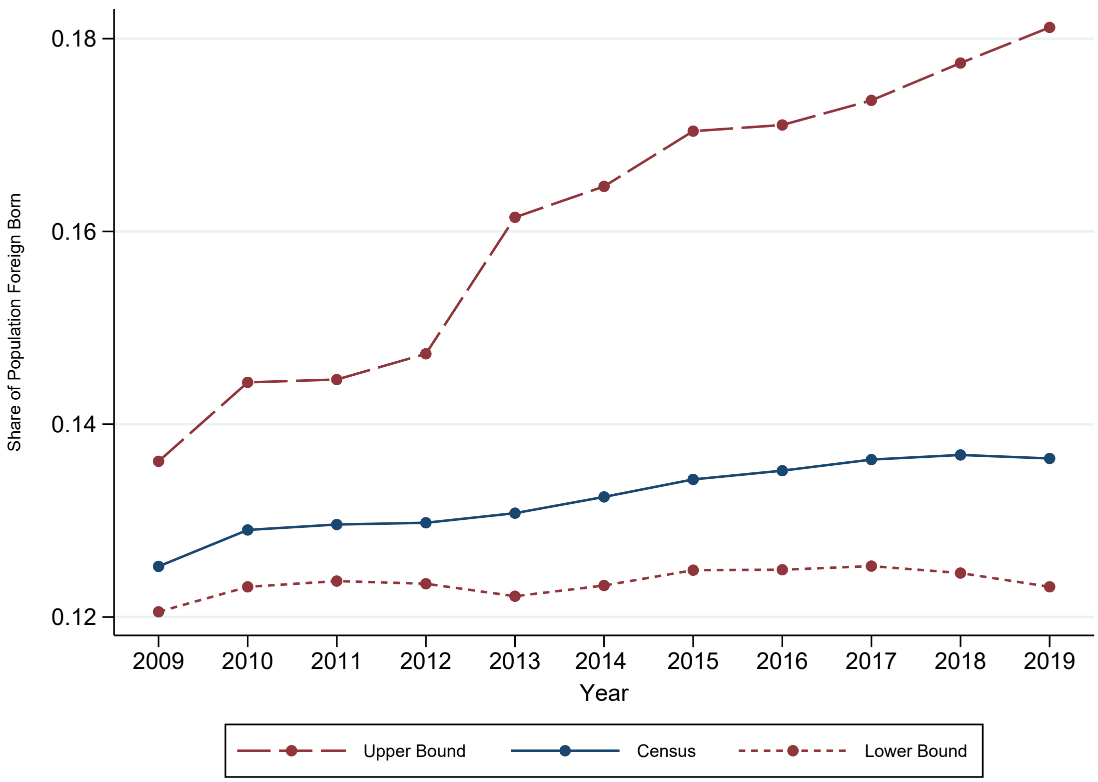
(B) Responded by Internet



(C) Responded by CAPI/CATI

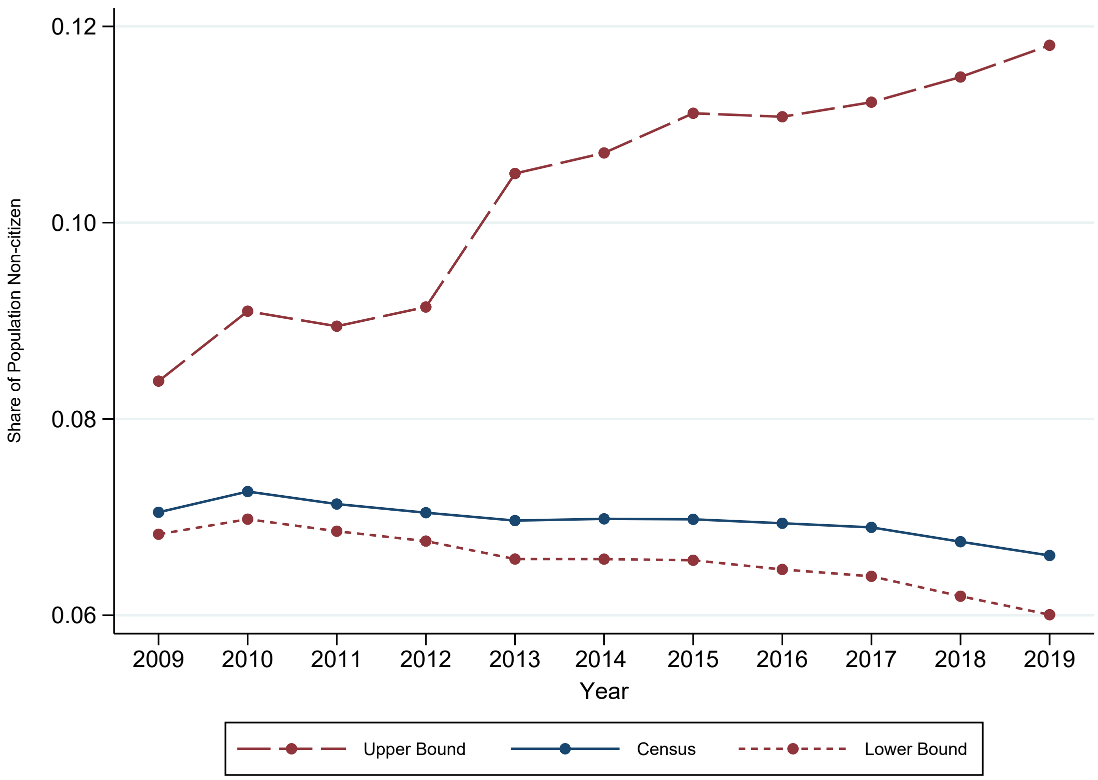
Notes: Author's own calculations using the American Community Survey. Shares are calculated using census person weights.

Figure 3.16: Interval Estimates of the Share of Population that is Foreign-Born



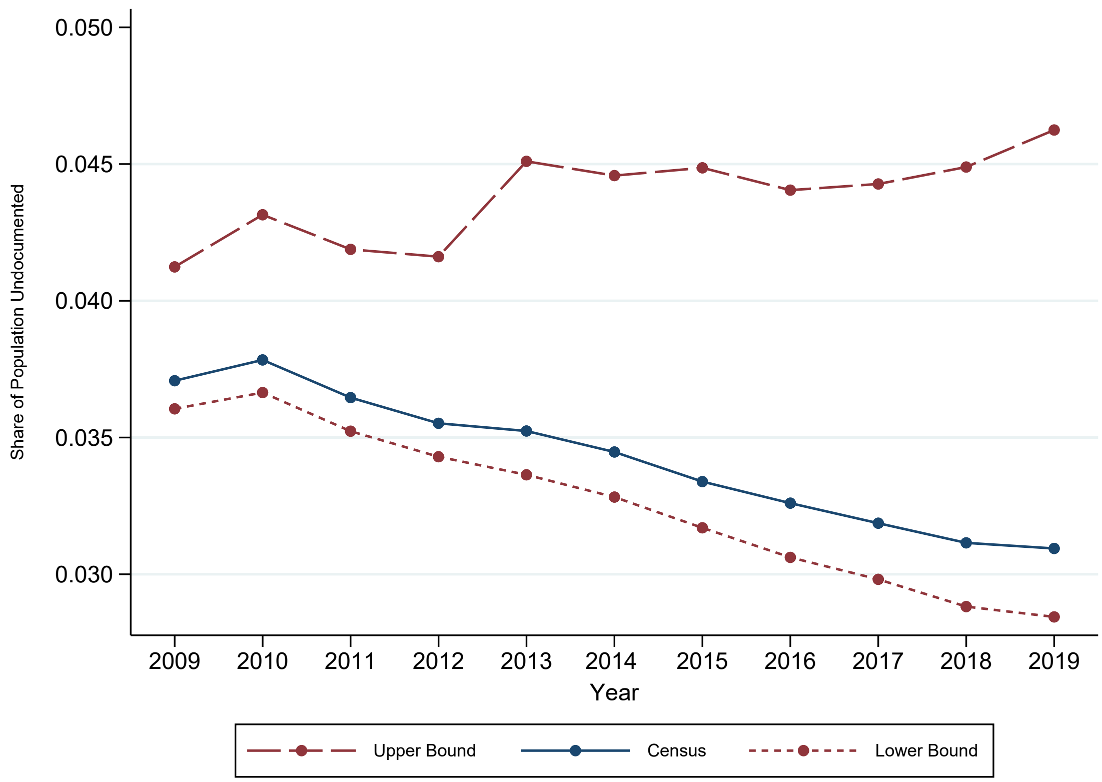
Notes: The long-dash line represents the upper bound of the share of foreign-born in the population where all imputed values are assigned as foreign-born. The short-dash line represents the lower bound of the share of foreign-born in the population where all imputed values are assigned as native-born. The solid line represents the share of foreign-born in the population using the imputed values from the Census under the assumption that nonresponse is conditionally random. Estimates are weighted using Census person weights.

Figure 3.17: Credible Interval Estimates of the Share of Population that is Non-citizen



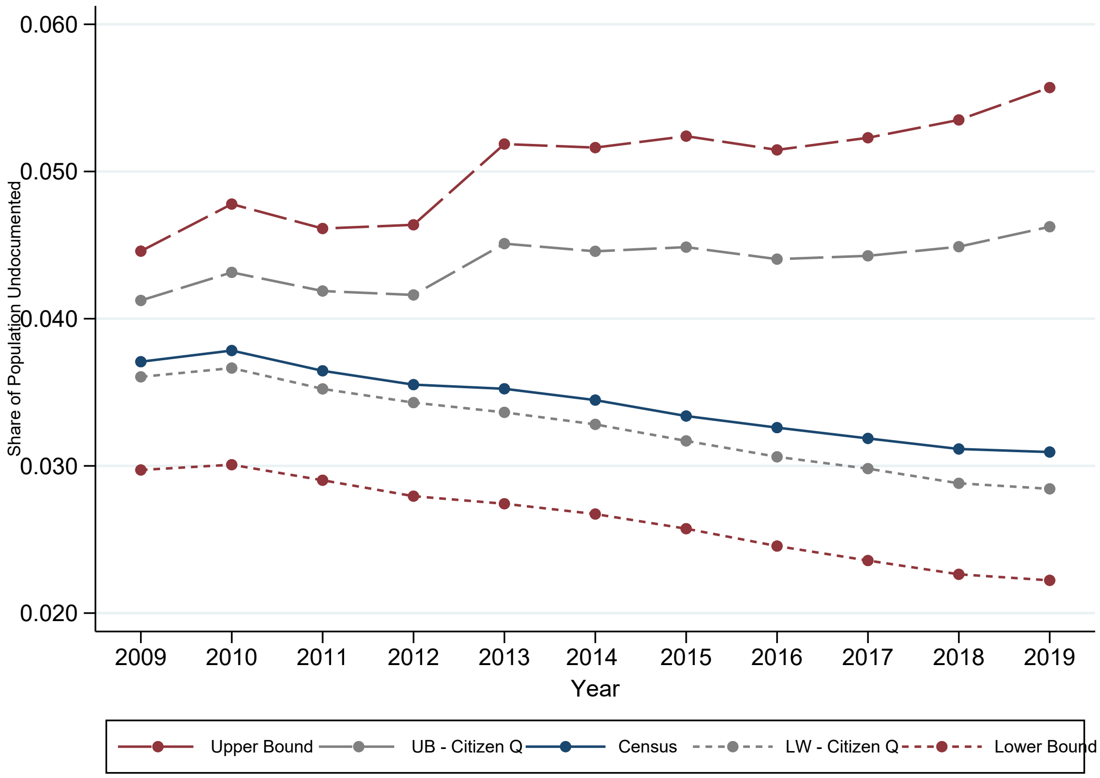
Notes: The long-dash line represents the upper bound of the share of non-citizens in the population where all imputed values are assigned as non-citizens. The short-dash line represents the lower bound of the share of non-citizens in the population where all imputed values are assigned as native-born. The solid line represents the share of non-citizens in the population using the imputed values from the Census under the assumption that nonresponse is conditionally random. Estimates are weighted using Census person weights.

Figure 3.18: Credible Interval Estimates of the Share of Population that is undocumented



Notes: The long-dash line represents the upper bound of the share of undocumented immigrants in the population where all imputed values are assigned as non-citizens. The short-dash line represents the lower bound of the share of non-citizens in the population where all imputed values are assigned as native-born. The solid line represents the share of non-citizens in the population using the imputed values from the Census under the assumption that nonresponse is conditionally random. After assignment the residual method procedure is conducted to create the estimates at each bound. Estimates are weighted using Census person weights.

Figure 3.19: Credible Interval Estimates of the Share of Population that is undocumented



Notes: The long-dash line represents the upper bound of the share of undocumented immigrants in the population where all imputed values are assigned as non-citizens. The short-dash line represents the lower bound of the share of non-citizens in the population where all imputed values are assigned as native-born. The solid line represents the share of non-citizens in the population using the imputed values from the Census under the assumption that nonresponse is conditionally random. After assignment the residual method procedure is conducted to create the estimates at each bound. Estimates are weighted using Census person weights.

Chapter 4 Imputing Treatment: Misclassification Bias from Item-Nonresponse in the Effects of DACA

4.1 Introduction

By 2018, the Deferred Action for Childhood Arrivals (DACA) program provided 824,000 undocumented youth temporary relief from deportation and work authorization. Enacted through executive memorandum in 2012, these benefits, subject to renewal every two years, are available to undocumented youth who are in school or had completed high school, and met the age and year of arrival criteria. The program was terminated in September 2017 by then President Trump and was later restored by President Biden in 2021. This chapter examines the degree of misclassification bias arising from item-nonresponse in the estimated effects of DACA eligibility on labor market outcomes.

A growing literature has developed estimating the effects of DACA eligibility on various outcomes. Previous work has shown that DACA improves labor market outcomes (Pope, 2016; Amuedo-Dorantes and Antman, 2017), improves health outcomes among children and adults (Venkataramani et al., 2017; Hainmueller et al., 2017; Giuntella and Lonsky, 2020), reduced teenage pregnancy (Kuka et al., 2019), reduces the propensity to commit crime (Gunadi, 2019). The effect of DACA eligibility on education has been mixed. Pope (2016), Amuedo-Dorantes and Antman (2017), and Hsin and Ortega (2018) show negative or insignificant effect on school attendance for those that are immediately eligible while Kuka et al. (2020) and Ballis et al. (2020) find positive relationship among those that may want to become eligible by completing high school. However, this literature has not taken into account the effects of non-classical measurement error arising from item-nonresponse.

The role of Census imputation on the effects of DACA has been largely ignored in the literature. Yet, item nonresponse among non-citizens, especially for the variables used to assign DACA eligiblity status, are significant and have become more severe over time. Assigning DACA eligibility requires information on a number of demographic factors; age, education level, year of immigration, and citizenship status. Nonresponse to the demographic questions used to assign DACA eligibility in the American Community Survey (ACS) averages 15% across the sample between the years 2005 to 2018. The significant degree of imputed values in the treatment variable (DACA eligiblity) has potential to lead to misclassification bias in the estimated effects of DACA eligiblity on labor market outcomes.

An existing literature has studied the consequences of misclassification in a binary treatment variable. Aigner (1973) shows OLS estimates are biased downward when there is misclassification in an exogenous binary regressor. More recently, Lewbel (2007) reaches the same attenuation bias from exogenous misclassification of a binary regressor in nonparametric and semiparametric regressions. While Aigner (1973) and Lewbel (2007) focused on exogenous misclassification, Nguimkeu et al. (2019) extend the literature to show that, when facing endogenous misreporting, there will be conditions in which sign reversal can occur.

The ACS uses a 'hot-deck' imputation procedure to assign a value to non-respondents. That is, non-respondents are assigned a value of a respondent who have similar characteristics. As discussed in Chapter 3, this procedure assumes that nonresponse is missing at random (MAR). When the MAR assumption holds for the questions used to assign DACA eligiblity, there is less concern of bias. Though, as also shown in Chapter 3, there is strong evidence that MAR assumption fails for the citizenship question (Brown et al., 2018). If the MAR assumption fails, this will lead to traditional misclassification bias which will attenuate the estimated effects.

I follow one approach to adjust for missclassification bias by removing non-respondents from the estimating sample. Using this approach leads to estimates of the labor market effects of DACA that are 22% to 77% higher than when including non-respondents, depending on the labor market outcome of interest. The results are robust to reweighting the respondent sample to take into account observed changes in sample composition. Nonresponse has been increasing over time, and therefore the likelihood of misclassification bias. This has an impact on the dynamic effects of DACA with the differences in magnitude between the respondent and the whole sample increases over time, reaching its peak in the year 2018. While the magnitude of the effects are larger than the whole sample, the effects of DACA become insignificant at the end of the sample. DACA, as a temporary program, has temporary effects.

The primary focus of this chapter is on the consequences of item nonresponse in the characteristics used in assigning DACA eligiblity. There is also potential for bias from item nonresponse in the outcomes of interest. Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) demonstrate that the OLS coefficients will be attenuated when using imputed outcomes if the imputation procedure does not use the treatment as a match criteria to allocate values. This type of bias has been commonly called match bias. When the imputation procedure does not condition on the treatment, the imputed outcomes will be uncorrelated with the treatment variable leading to attenuated estimated. Match bias will occur even if data is missing at random.

As DACA eligiblity is not a match criteria for any of the outcomes of interest, match bias will attenuate the estimated effects of DACA on labor market outcomes. In this chapter, I show that nonresponse to the variables used to assign treatment are highly correlated with nonresponse to the outcomes of interest. Nearly all of the individuals with imputed outcomes have at least one of the variables used for assigning DACA imputed as well. Thus, adjusting for misclassification bias through dropping item non-respondents also adjusts for the potential match bias from having imputed outcomes.

I also propose an alternative method to deal with bias arising from item nonresponse. I produce bounding estimates by assigning all item non-respondents as either DACA-eligible or DACA ineligible. Misclassification can be caused by true DACA-eligible non-citizens being imputed as DACA-ineligible (false negatives) or vise versa (false positives). By using bounds for the imputed values, each sample is composed of only one type of misclassification. This procedure is similar to that recommended in Chapter 3 to take into account uncertainty caused by item nonresponse. In the sample in which I assign all non-respondents as DACA ineligible, the effects are similar, to the respondent only sample though slightly larger in magnitude. When assigning all non-respondents as DACA eligible, the effects are similar to those of the whole sample but slightly larger. The results are a strong indication that false positives (DACA-ineligible assigned as DACA-eligible) are the predominant source of misclassification.

Whichever method used to deal with bias caused from item nonresponse, they are all preferred than when using the imputed values. As nonresponse is of greater significance among non-citizens, researchers analyzing immigration policy should be cautious of using imputed data when treatment is based on a function of demographic characteristics.

The paper is structured as follows: Section 4.2 details the data used for the analysis; Section 4.3 describes the model specification used; Section 4.4 provides the results of DACA when adjusting for misclassification and match bias; Section 4.5 shows the results from an event study framework; Estimates using a bounds analysis are presented in Section 4.6; while Section 4.7 concludes.

4.2 Data and Variable Construction

To analyze the effect of DACA eligiblity on labor market outcomes, I use data from the IPUMS ACS (Ruggles et al., 2020). The sample starts with the year 2005 since it's the first year with a full one-percent sample of the U.S. population. Sample year 2018 is the last year for which the survey data is available. ACS contains a detailed set of standard socio-demographic characteristics and labor market outcomes (e.g. employment, labor force participation, annual income). The baseline sample comprises of all non-citizens ages 18 to 35 with at least a high school degree (or equivalent).

The ACS also contains information on US citizenship status, number of years spent in the US, quarter of birth, and educational attainment, which can be used to determine respondents' DACA eligibility status. DACA-eligible individuals are defined as those who: (1) were under the age of 31 as of June 15, 2012; (2) have lived in the U.S. since June 15, 2007; (3) entered U.S. before reaching 16th birthday; (4) have at least a high school degree (or equivalent); (5) were born outside the U.S. (orits territories); and (6) are not U.S. citizens.[1] Similar to Pope (2016) and Giuntella and Lonsky (2018), I am estimating the effects of DACA on those who are immedietly eligible for the program.

The ACS does not have data on legal status nor criminal history. This lack of information will lead to misclassification error and attenuation bias of the estimated effects.

[1] To define the DACA-eligible population in year 2012 and before, the criteria are restricted to non-citizens who were: (1) Under the age of 31 as of June 15 of the previous calendar year; (2) Arrived to the U.S. prior to their 16th birthday, (3) Has lived in the U.S. for at least 6 years, (4) has at least a high school degree (or equivalent); (5) were born outside the U.S.; and (6) are not U.S. Citizens.

Though it is an importance source of bias, this is not the focus of the paper. The following sections show that attenuation bias derived from imputation of the variables used to assign DACA eligiblity is of similar or greater magnitude.

4.2.1 Item Nonresponse in the ACS

The responses to four questions are used to assign DACA eligiblity; (1) educational attainment, (2) age, (3) year immigrated, and (4) citizenship status. Figure 4.1 plots the share of item non-response for each question used for assigning DACA eligiblity. A considerable share of the non-citizen sample did not respond to these questions. The year immigrated question has the highest nonresponse rate at 10%, on average. Education has the second highest average non-response at 8% of the sample. From Figure 4.1, a clear break in the trend in item non-response occurred in 2012 for the educational attainment and immigration related variables where nonresponse rose rapidly. Refusal to respond to the citizenship question more than doubled from before 2012. The other variables saw an increase in non-response rates by about 50%. The nonresponse rate in age doubled in the year 2008 but has stayed stable since.

The trend break that occurred in 2012 is likely driven by changes in the survey collection methodology and a reduction in the number of failed-edit follow up calls due to budgetary reasons (Clark, 2014). The change in survey collection is the introduction of an internet response mode (Clark, 2014). The methodological changes in 2013 did not have an impact on the response rate of other demographic questions such as race, Hispanic origin, sex, age, nor housing tenure questions (O'Hare, 2018; Clark, 2014). These questions are asked before the citizenship question so ordering might be an issue on item nonresponse when including sensitive questions on surveys such as citizenship status (O'Hare, 2018).

The degree of nonresponse to each individual question may not appear as severe, and thus so to the concern for misclassification bias. The possible issues associated with item nonresponse, though, compounds when treatment is a function of multiple demographic variables. Figure 4.2 shows the magnitude of the nonresponse rates for the sample used to estimate the labor market effects of DACA. This figure plots the share of immigrants that had DACA eligibility assigned with at least one imputed value by treatment status over time. The solid line shows the overall share of immigrants in the sample that had treatment imputed. The long dash line shows the share of immigrants assigned as DACA-eligible that had at least one of the key variables imputed. The short dash line shows the share of immigrants assigned as DACA-ineligible that had at least one of the key variables imputed. Imputation of treatment status is substantial and has doubled throughout the sample period. On average, DACA-eligibility is imputed for 15% of the sample and increased to almost 20% by 2018. Those assigned as DACA-eligible have considerably higher imputation at an average of 18.2%. By 2018, almost 25% of the assigned DACA-eligible group had their treatment status imputed. For those assigned DACA-ineligible, about 17% of the control group had their status imputed by 2018. The higher rates of nonresponse among the DACA-eligible may be driven by the fact DACA-eligible individuals are younger than the DACA-ineligible control group. I show in chapter 3 that nonresponse to the citizenship question is higher among younger immigrants compared to older immigrants. If there is an error in the imputation method and MAR fails, misclassification error will occur and bias the estimated results.

4.3 Empirical Strategy

To estimate the effects of DACA and quantify the degree of match and misclassification bias, I follow a modified version of the difference-in-differences approach proposed by Pope (2016) and used by many other researchers estimating the labor market effects of DACA. This model compares DACA eligible non-citizens to DACA ineligible non-citizens before and after the implementation of DACA in 2012. The baseline (whole) sample comprises of all non-citizens ages 18 to 35 with at least a high school degree (or equivalent). The regression estimated is as follows:

$$\begin{aligned} Y_{ist} = \beta_0 + \beta_1(eligible_{it} \cdot post_t) + \beta_2 eligible_{it} + \beta_3 post_t \\ + \beta_4 X_{it} + \beta_5 W_{it} + \beta_6 Z_{st} + \gamma_t + \gamma_s + \gamma_s t + \epsilon_{ist} \end{aligned} \tag{4.1}$$

Where Y_{ist} is the labor market variable of interest (e.g., working, unemployed, in the labor force, etc.) for immigrant i in state s in year t. The variable $eligible_{it}$ is an indicator if the individual meets all observable DACA requirements. The variable $post_t$ is a binary variable equal to one if the year is after 2012, and equal to zero if the year is from 2005 to 2012. The vector X_{it} contains demographic controls including education, sex, race, ethnicity, marital status. The vector W_{it} non-parametrically controls for the eligibility criteria by including fixed effects for individual $i's$ age and age when they arrived in the United States. Z_{st} is a vector of state-by-year controls including the state unemployment rate, state minimum wage rate, as well as indicators if the state has universal E-Verify laws, public E-Verify laws, if the state grant in-state tuition to undocumented immigrants, and whether the state provides undocumented immigrants access to driver's license. The vectors γ_t and γ_s allow for time and state fixed effects, respectively. Lastly, $\gamma_s t$ allows for state-specific time trends. When estimating equation 4.1, standard errors are clustered at the state-by-year level.

The parameter of interest, β_1, is the coefficient on the interaction term between $eligible_{it}$ and $post_t$. The coefficient, β_1, therefore corresponds to the intent-to-treat effect of DACA relative to ineligible non-citizens.

To be exact, as I use non-citizenship as a proxy for being undocumented, the coefficient β_1 is an attenuated measure of the intent-to-treat effects of DACA. Estimates from the DHS indicate the share of unauthorized immigrants among the non-citizen population aged 18-35 is around 61% (Baker and Rytina, 2013). Attenuation bias from misclassifying authorized immigrants as DACA-eligible is expected to be around 60% (Pope, 2016). The true intent-to-treat estimates will be 1.6 times larger than the difference-in-differences estimates produced using non-citizens as a proxy for unauthorized status.

The results will further be attenuated due to misclassification bias from using imputed values in the DACA eligiblity indicator when MAR assumption fails. With an average of 15% of the sample having DACA eligiblity assigned using imputed values, there is the possibility of considerable attenuation bias. To quantify the degree of misclassification bias from item nonresponse to the questions used to assign DACA eligiblity and the outcomes of interest, I estimate Equation 4.1 using various sub-samples from the whole sample based on the non-citizens response status to the key questions. Section 4.4 provides more detail on each sub-sample.

4.4 Results

4.4.1 Effects of DACA Adjusting for Misclassification Bias

Misclassification in the DACA Eligibility indicator caused by the Census imputation procedure will bias the estimated effects of DACA. A simple method to adjust for misclassification bias caused from item nonresponse is to drop those whose DACA eligibility is assigned using imputed values (Bollinger and Hirsch, 2006). I estimate equation 4.1 using the whole sample, a respondent only sample, and a non-respondent only sample separately. A respondent is defined as an individual who responded to all questions used to assign eligibility. A non-respondent is defined as an individual who did not respond to at least one question used in assigning eligibility.

The estimates from equation 4.1 using the whole sample (Column 1), respondent sample (Column 2), and non-respondent sample (Column 3) are shown in Table 4.1. The results using the whole sample are similar to those obtained in Pope (2016).[2] The results using the whole sample will be attenuated towards zero when including noncitizens whose eligibility status may be misclassified due to the imputation procedure. Column 4 and Column 5 of Table 4.1 show the difference in magnitude between the respondent sample and the whole sample and the non-respondent sample and the whole sample, respectively.

There is a sizable differences in the estimated magnitude of the effects of DACA Eligibility between the respondent sample and the whole sample. This is suggestive of severe misclassification bias when including non-respondents in the sample. Among the outcomes in which DACA eligibility has a significant effect, the difference in magnitude range from 22.4% (unemployment) to as high as 77.9% (total income). For each outcome of interest, the magnitude of the differences in the estimated effects is at least as large as the share of the sample that is has eligibility status imputed (15.4% across all years in the sample).

While there is no statistically significant difference between both sample estimates, the differences in the magnitude of the coefficients of interest is economically significant. The effect of DACA on labor force participation and on the probability of working is about 1 p.p.

[2] The differences are driven by the inclusion of more years and additional state-year immigration policy controls.

and 1.2 p.p larger in the sample with no imputed treatment compared to the whole sample estimates. For labor force participation, the effects of DACA are 52% larger than when using the whole sample. The effects of DACA on the likelihood of working are 39% larger in the sample adjusting for misclassification bias. Economically, these are large differences. With 1.7 million estimated DACA eligible individuals (Passel and Lopez, 2012), a 1.2 p.p. difference in the effects of DACA on the likelihood of working amounts to an additional 20,000 individuals being employed due to DACA. For total income, the effects of DACA-eligibility are 22.7% in the respondent sample compared to 12.7% in the whole sample. For both self-employment and school attendance, the effects in both samples are still both economically and statistically insignificant.

For most outcomes of interest, the effect of DACA eligibility among the sample in which eligibility assigned using imputed values (Column 3) is close to zero. This is what would be expected if nonresponse is random. For usual hours of work and total income, the effects of DACA-eligibility are of opposite sign. Only school attendance shows a positive and significant 1.5 p.p. effect of DACA eligibility among the non-respondent sample.

An important source of misclassification error in assigning treatment status for DACA eligibility is due to lack of information on legal status and on criminal history. It is estimated that 40% of non-citizens ages 18 to 35 are authorized immigrants (Baker and Rytina, 2013). Attenuation bias from misclassifying authorized immigrants as DACA-eligible is expected to be around 60% (Pope, 2016). That is, the true intent-to-treat estimates will be 1.6 times larger than the difference-in-differences estimates produced using non-citizens as a proxy for unauthorized status. Here, we see that misclassification bias from using imputed characteristics to assign eligibility status is just as large of an issue in the difference-in-differences estimates as what is caused from not being able to distinguish between authorized and unauthorized immigrants among the non-citizen population.

Adjusting for Selection on Observables

In Table 4.1, I dropped individuals in which DACA eligibility assignment used at least one imputed value to adjust for misclassification bias. If nonresponse to the questions used for assigning DACA eligibility status is random, then the composition of the respondent and whole sample will be the same. If this is not the case and nonresponse is not random, then the estimates can differ from compositional changes alone.

In table 4.2, I estimate the effects of DACA eligibility on the respondent sample using inverse probability weighting (IPW) to correct for selection on observables as done in Bollinger and Hirsch (2006). To account for the compositional changes correlated with observables, I first run a probit equation with response to all questions used in the assignment of DACA eligiblity as a binary dependent variable. The controls used are the same as in Equation 4.1. This estimates a probability of responding to all questions conditional on the observed characteristics. I then weight the respondent sample by the inverse probability of response. In short, the IPW sample reweights the respondent sample to be representative of the whole sample. Reweighting does not correct for possible selection on unobservables (unobserved factors correlated with the outcomes of interest but not correlated with the independent variables).

Table 4.2 Column 5 shows the ratio between the IPW respondent sample and the unweighted respondent sample. There is virtually no difference between the estimates once controlling for self-selection in observables. The largest difference is 2.2% in the estimated effects of DACA eligibility on total income. Differences in the estimates for school attendance and self-employment are driven by differences in values at the 4th decimal place.

Adjusting for Each Question Separately

The responses to four questions are used to assign DACA eligibility status. To see if a specific question is driving the differences between the respondent and the whole sample, I next estimate Equation 4.1 by dropping non-respondents to each question individually and are displayed in Table 4.3. Column 1 of Table 4.3 drops non-respondents to the education question, Column 2 drops non-respondents to the age question, Column 3 drops non-respondents to the year immigrated question, while Column 4 drops non-respondents to the citizenship question. Column 5 of Table 4.3 drops all who have at least one question imputed and is the same as Column 2 of Table 4.1. Year immigrated has the highest share imputed at 10.2% of the sample, followed by education (8.5% of the sample), citizenship (5.7%), and age at 1.6%. As expected, the sample that drops non-respondents to the year immigrated is closest to the estimates when dropping all with at least one imputed response while the sample dropping only those who did not respond to the age question is closet to the results from the whole sample.

It can be seen from Table 4.4 that, in the sample which responded to the age question, 13% of the sample still had at least one other outcome used for assigning DACA eligibility imputed. Thus, it is still likely to contain misclassification error from the imputation procedure. In the sample in individuals who responded to the year immigrated question, only 4% of the sample had at least one other question (mainly education attainment) imputed and therefore less of the sample where these individuals are dropped will face misclassification bias from the imputation procedure.

4.4.2 Effects of DACA Adjusting for Match Bias

In the prior subsection, I dealt with misclassification bias in DACA eligibility caused by item nonresponse. The estimates will also be biased due to nonresponse to the questions of interest. This match bias is caused by the fact the Census does not use DACA eligibility as a match criteria when matching donor answers to non-respondents (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2006). As the outcomes of citizens and other non DACA-eligible immigrants are more likely to be used as as an assigned value for DACA-eligible non-citizens, there is significant potential for match bias.

Match bias can be described as misclassification bias (Hirsch and Schumacher, 2004). The imputed value to the outcome of interest can be treated as a valid observation whose DACA eligiblity status is misclassified. This misclassification will lead to the standard misclassification bias discussed earlier. As will be shown, match bias and misclassification bias in this sample are strongly linked. Noncitizens that do not respond to the DACA eligiblity questions are highly likely to not respond to the outcomes of interest as well.

To deal with the potential for match bias from nonresponse to the outcomes of interest, I now define a respondent as an individual that responded to all questions used in assigning DACA eligibility and also to the respective outcome of interest. Column 3 of Table 4.5, shows the estimated effects using this new respondent sample. Column 5 of Table 4.5 show the ratio between the old respondent sample and the respondent sample excluding imputed outcomes of interest. The results are slightly larger when dropping the imputed outcome values. The largest difference is in the effects of DACA eligibility on total income at 6.2%

larger magnitude when dropping those who did not respond to the income question.

Table 4.6 shows why the results are very similar. The correlation with not responding to at least one of the demographic questions used to assign DACA eligibility and not responding to labor market questions is very high. Among the sample that had DACA eligibility imputed (Column 1 of Table 4.6), nonresponse is as low as 33% for the class of worker question to as high as 42% for total income question. Among those that responded to all questions used to assign treatment (Column 2), nonresponse to the labor market questions ranged from 1% (employment and school attendance) to at most 5% (total income).

Essentially, adjusting for misclassification bias by dropping non-respondents to the questions used to assign eligibility adjusts for almost all of the potential match bias from nonresponse to the outcomes of interest. Even if there is no misclassification bias from imputation in the treatment variable and the imputation procedure perfectly assigns the demographic characteristics, as DACA eligibility status is not a match criteria for any of the outcomes of interest there is a strong possibility of match bias from nonresponse in the outcomes of interest. As citizens, whether native-born or naturalized, make up the vast share of respondents, DACA-eligible non-citizens or ineligible non-citizens that do not respond to the labor market questions are likely to be assigned an outcome of that of a similar citizen. Assuming there is no effect on DACA and there are no differential trends between citizens and non-citizens, the results will be attenuated towards zero. If there is differential trends in the outcomes of interest between citizens and non-citizens the bias will depend on the degree of difference between the trends. This may explain why the effects of DACA among the non-respondent sample in Table 4.1 are significant for usual hours of worked, school, and total income.

4.4.3 Effects of DACA on Response Rates

A possible concern of using the respondent only sample is that DACA affects the response rates directly. That is, the enactment of DACA lead to DACA eligible individuals to be more likely to respond to the ACS. I now move to estimating the effects of DACA on response rates. I first focus on the likelihood of nonresponse to each question used to assign DACA eligiblity using the whole sample. An important limitation needs to be stated. Since there is misclassification bias, the results will be attenuated towards zero. As such, this is only suggestive evidence of the effect of DACA on response rates.

The results of the effects of DACA on the response rates to the demographic questions using the whole sample are presented in Table 4.7. There is no effect of DACA eligibility on the response rates to the education question, age question, or year immigrated question. DACA eligible non-citizens see a significant 0.7 p.p. (18%) drop in the probability of not responding to the citizenship question after the enactment of DACA relative to ineligible non-citizens. Overall, there is no change in the probability a DACA-eligible individual has their Eligible status assigned using imputed values compared to ineligible non-citizens (Column 5).

If the imputation procedures used perfectly assigns values and there is no misclassification error, the estimates in Table 4.7 show there is no significant impact of DACA on eligible likelihood of responding to at least one of the four questions used to assign treatment. This is important as it provides evidence that the results presented above are not being driven by changes in who is responding to the survey after the implementation of DACA. Though as mentioned, this is only suggestive evidence that DACA does not impact survey question response rates.

Table 4.8 shows the effect of DACA eligibility on the likelihood of not responding to the questions used for the outcomes of interest. Row 1 of Table 4.8 shows the effect of DACA on non-response to the labor market questions using the whole sample while Row 3 uses a sample of those who responded to all questions used to assign eligibility. In the whole sample, DACA-eligible non-citizens are more likely to not respond to the class of worker question and total income question by 8.5% and 5.7% relative to ineligible non-citizens, respectively. In the whole sample, DACA-eligible non-citizens are 7.9% more likely to respond to the question on school attendance. Among the sample that responded to all questions used to assign treatment (to reduce the potential for misclassification bias), the likelihood of nonresponse to the labor market questions is even greater among DACA eligible non-citizens after the enactment of DACA relative to ineligible non-citizens. Among the respondent sample, after DACA, non-response to the class of worker question went up 32.3%, up 14.3% for school attendance question, and up 14.6% for total income.

This is a surprising result. After DACA, DACA-eligible non-citizens have become more difficult to reach than those of their ineligible non-citizen counterparts. Column 6 of Table 4.8 estimates the effect of DACA eligibility on an alternative measure of difficulty-of-reaching a respondent; whether they responded through phone or in-person interview. If a household does not respond by mail or internet within the first month, the Census calls the household for an interview. If that fails, in the third month the Census visits the household for an in person interview. After DACA, DACA-eligible non-citizens where 1.5 p.p, (2.1%) more likely to respond through phone or in-person interview. This provides further evidence that DACA-eligible non-citizens are becoming harder to reach for surveys.

A major concern on using a respondent only sample is that the largest effects may be driven by self-selection. That is, those who are most likely to benefit the most are more likely to respond to the survey. Due to the elimination of deportation fears and work authorization, it would be expected that item-nonresponse rates would decrease among the DACA-eligible population after the enactment of DACA. But as show in in Table 4.8, that is not generally the case.

4.5 Dynamic Effects of DACA Adjusting for Misclassification Bias

As shown in Figures 4.2 and 4.2, item-nonresponse has been increasing over time. It is therefore expected that treatment imputation will affect the estimated dynamic effects of DACA eligibility. I modify Equation 4.1 into an event study framework described below.

$$y_{ist} = \sum_{j \neq t^*-1} \delta_j \cdot year_{j=t} \cdot eligible_{ist} + \sum_{j} \beta_j \cdot year_{j=t} + \alpha Eligible_{ist} + \epsilon_{ict} \tag{4.2}$$

Rather than interact an indicator for being DACA eligible with a post DACA indicator, I interact the DACA eligible indicator with a series of year dummies. Year 2012 is assigned as the reference year. All controls are the same as Equation 4.1. Standard errors are clustered at the state-year level.

Figure 4.3 and Figure 4.4 shows the results of estimating the event study model of equation 4.2 using the whole sample (solid line) and a sample with no imputed treatment (dash line). The figure plots the coefficients of the interaction of the eligible indicator with year dummies. The reference (omitted) year is 2012. The difference between the whole sample and the respondent only sample fans out over time as expected. As with the estimated effects in the difference-in-differences model, the respondent sample shows the effects of DACA to be of larger magnitude than the whole sample for all significant outcomes of interest. Only self-employment and school attendance do we see almost exact effects when using the whole sample or respondent sample. Whether using the respondent only sample or the whole sample with individuals with imputed treatment, the effects of DACA begin to decrease rapidly and become insignificant by the year 2018. The effects of a temporary program are, in fact, temporary.

Figure 4.5 re-estimates the results of Table 4.7 in an event study framework. The results are similar to Table 4.7, there is no significant effect of DACA on the response rates to the questions used to assign treatment except for the citizenship question. Although, the effects of DACA eligiblity to responding to the citizenship status question are only statistically significant in the year 2016. Figure 4.6 re-estimates the results from Row 1 of Table 4.8 in an event study framework. Here, the results of Row of Table 4.8 are not robust. DACA does not appear to cause differential trends in the response rate to the outcomes of interest after 2012.

4.6 Alternative Method to Dealing with Non-response

The above estimates have shown misclassification bias caused by item-nonresponse is a major source of bias in the estimated effects of DACA when including non-respondents in the sample. Among the sample that has individuals with imputed DACA eligibility status, the rate of misclassification is not known to researchers. Misclassification can be caused by true DACA-eligible non-citizens being imputed as DACA-ineligible (false negatives) or vise versa (false positives).

Rather then dropping non-respondents as is done in Section 4.4, I now perform an alternative method to deal with item non-response in the spirit of Manski (2016). As eligibility may never be known for the non-respondents without access to administrative records, rather then drop those whose DACA eligibility status was assigned using imputed values, I impute the extreme cases. In one sample all who have at least one question imputed are assigned as DACA-eligible. In this sample, only false positives will be in the imputed sample. In the second sample, all who have at least one question imputed are assigned as DACA-ineligible. In this sample, only false negatives will be in the imputed sample. While we may not know the degree of misclassification, the estimated effects will provide suggestive evidence which type of misclassification is of bigger concern. This is the method proposed in Chapter 3 as a way to take into account uncertainty caused by item nonresponse.

I estimate Equation 4.1 using both the samples detailed above. The results are presented in Table 4.9. Column 2 of Table 4.9 shows the estimated effects of DACA when all non-respondents are treated as DACA-eligible. Column 3 of Table 4.9 shows the estimated effects of DACA when all non-respondents are treated as DACA-ineligible. When we treat all non-respondents as DACA-ineligible, the estimated effects are very similar to the respondent only sample but slightly lower in magnitude for all statistically significant outcomes. When we treat all non-respondents as DACA-eligible, the estimated effects are much smaller then the respondent only sample for all significant outcomes. Importantly, when treating all non-respondents as DACA-eligible, the estimated effects of DACA are all larger than when using a sample that includes the imputed values. The results are a strong indication that false positives (DACA-ineligible assigned as DACA-eligible) are the predominant source of misclassification.

4.7 Conclusion

Nonresponse in the main surveys used by researchers has been increasing rapidly over the years. Nonresponse rates among the non-citizen population are even greater than that of the general population. When treatment is assigned using demographic questions, as is typically the case, the concern for misclassification bias arises. This leads to significant potential for misclassification bias in the estimated effects of immigration policy on this population.

In this paper, I have shown that item nonresponse to the demographic variables used to assign DACA eligiblity and on the outcomes of interest in the ACS is significant among the non-citizen population. Over 15% of the sample has their DACA eligiblity status assigned using at least one imputed value. The share of the sample where DACA eligiblity is assigned using imputed values doubled during the sample period of 2005 to 2018.

This leads to considerable misclassification bias in the estimated effects of DACA eligiblity on labor market outcomes, with the magnitude of the bias being typically greater than the share of non-respondents. The results are consistent when re-weighting the respondent sample by the inverse of the probability of responding to take into account compositional changes. As nonresponse to the demographic variables are highly correlated with nonresponse to the outcome variables, adjusting for misclassification bias adjusts for almost all of the potential match bias as well. As nonresponse, and therefore the likelihood of match and missclassification bias, has increased over the years, the dynamic effects of DACA are impacted. The difference in the effects of DACA between the whole sample and the respondent only sample are greatest in the final years of the sample when nonresponse is greatest.

This chapter focused on misclassification bias from item nonresponse affecting the intent-to-treat estimates of the effects of DACA eligiblity on labor market outcomes. It is also important to note that the estimated treatment-on-the-treated effects estimated in Chapter 2 will also be impacted by this misclassification bias. To assign DACA participation in Chapter 2, I used DACA eligiblity and country-of-origin. As such, the results here indicate that the estimated treatment-on-the-treated effects documented in Chapter 2 are likely attenuated towards zero.

DACA did not have an effect of the likelihood of having their DACA-eligiblity status assigned using imputed values. Although, it had a positive effect on responding to the citizenship question specifically. I also find that DACA led to DACA-eligible individuals to become more likely to not respond to the outcomes of interest relative to non DACA eligible non-citizens. Though, these estimates are not robust to an event study specification.

I implement an alternative method to dealing with potential misclassification bias from item nonresponse by using bounds. I assign the extreme case scenarios to all non-respondents. This is similar to the method proposed in Chapter 3 to deal take into account uncertainty in estimates caused by item nonresponse. The benefit of the bounds procedure is that it provides extreme case scenarios of DACA eligibility. Whether assigning all non-respondents as DACA-eligible or as DACA-ineligible, the magnitude of effects are greater than when using the census imputed values in the whole sample. In the case of the effects of DACA, the bounds are between the estimates using the whole sample and the estimates using the respondent only sample. Whichever procedure used, whether dropping non-respondents, reweighing the respondent sample with inverse probability weights, or doing a bounding analysis, they are all preferred over using the whole sample with imputed values.

4.8 Chapter 4 Tables

Table 4.1: The Effects of DACA-Eligibility Adjusting for Non-Response

Variables	(1) Whole Sample	(2) Respondent Sample	(3) Non-Respondent Sample	(4) (2)/(1)	(5) (3)/(1)
Labor Force	0.019***	0.029***	-0.004	1.518	-0.197
	(0.005)	(0.005)	(0.008)		
Working	0.030***	0.042***	-0.005	1.386	-0.162
	(0.005)	(0.005)	(0.008)		
Unemployed	-0.016***	-0.020***	0.001	1.224	-0.081
	(0.003)	(0.004)	(0.006)		
Hours Worked	0.672**	1.127***	-0.601*	1.677	-0.894
	(0.261)	(0.274)	(0.327)		
Self-Employed	-0.004	-0.004	-0.005	1.000	1.486
	(0.002)	(0.003)	(0.004)		
School	0.005	0.000	0.015**	-0.021	3.326
	(0.005)	(0.006)	(0.007)		
Total Income (IHS)	0.127**	0.227***	-0.131*	1.779	-1.024
	(0.055)	(0.059)	(0.074)		

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.2: The Effects of DACA-Eligibility Adjusting for Non-Response using IP Weights

Variables	(1) Whole Sample	(2) Respondent Sample	(3) IPW Respondent	(4) (2)/(1)	(5) (3)/(2)
Labor Force	0.019***	0.029***	0.029***	1.518	1.003
	(0.005)	(0.005)	(0.005)		
Working	0.030***	0.042***	0.042***	1.386	1.010
	(0.005)	(0.005)	(0.005)		
Unemployed	-0.016***	-0.020***	-0.020***	1.224	1.010
	(0.003)	(0.004)	(0.004)		
Hours Worked	0.672**	1.127***	1.138***	1.677	1.010
	(0.261)	(0.274)	(0.274)		
Self-Employed	-0.004	-0.004	-0.004	1.000	1.257
	(0.002)	(0.003)	(0.003)		
School	0.005	0.000	-0.000	-0.021	-4.000
	(0.005)	(0.006)	(0.006)		
Total Income (IHS)	0.127**	0.227***	0.232***	1.779	1.022
	(0.055)	(0.059)	(0.058)		

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.3: The Effects of DACA-Eligibility Adjusting for Non-Response by Question

	Drop Nonresponse By question				
	(1)	(2)	(3)	(4)	(5)
Variables	Education	Age	Year Immigrated	Citizenship	At Least One
Labor Force	0.024***	0.020***	0.030***	0.024***	0.029***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Working	0.036***	0.031***	0.042***	0.035***	0.042***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Unemployed	-0.018***	-0.016***	-0.019***	-0.016***	-0.020***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Hours Worked	0.947***	0.683***	1.156***	0.899***	1.127***
	(0.265)	(0.262)	(0.266)	(0.260)	(0.274)
Self-Employed	-0.003	-0.004*	-0.004	-0.004*	-0.004
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
School	0.000	0.005	0.001	0.002	0.000
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)
Total Income (IHS)	0.179***	0.133**	0.230***	0.172***	0.227***
	(0.056)	(0.055)	(0.057)	(0.056)	(0.059)
Share Sample Dropped	0.085	0.016	0.102	0.057	0.141

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.4: Share of Question Imputed By Question Response

	Share of Question Imputed				
	(1)	(2)	(3)	(4)	(5)
Variables	Education	Age	Year Immigrated	Citizenship	At Least One
Education					
responded	0	0.01	0.05	0.02	0.06
imputed	1	0.07	0.70	0 .50	1
Age					
responded	0.08	0	0.10	0.05	0.13
imputed	0.40	1	0.47	0.27	1
Year Immigrated					
responded	0.03	0.01	0	0.01	0.04
imputed	0.58	0.07	1	0.49	1
Citizen					
responded	0.05	0.012	0.06	0	0 .09
imputed	0.74	0.08	0.88	1	1
At Least One Question					
responded (to all qs.)	0	0	0	0	0
imputed (At least one)	0.60	0.11	0.72	0.40	1

Standard errors in parenthesis
* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.5: The Effects of DACA-Eligibility Adjusting for Non-Response of Treatment and Outcome Questions

Variables	(1) Whole Sample	(2) Respondent Treatment Sample	(3) Respondent All Q Sample	(4) (2)/(1)	(5) (3)/(2)
Labor Force	0.019***	0.029***	0.029***	1.518	1.00
	(0.005)	(0.005)	(0.005)		
Working	0.030***	0.042***	0.042***	1.386	1.01
	(0.005)	(0.005)	(0.005)		
Unemployed	-0.016***	-0.020***	-0.020***	1.224	1.01
	(0.003)	(0.004)	(0.004)		
Hours Worked	0.672**	1.127***	1.148***	1.677	1.018
	(0.261)	(0.274)	(0.280)		
Self-Employed	-0.004	-0.004	-0.004	1.000	1.09
	(0.002)	(0.003)	(0.003)		
School	0.005	0.000	-0.001	-0.021	7.00
	(0.005)	(0.006)	(0.006)		
Total Income (IHS)	0.127**	0.227***	0.2409***	1.779	1.062
	(0.055)	(0.059)	(0.060)		

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.6: Share of Outcome with Imputed Value By Assignment Type

	(1) Imputed Treatment	(2) Non-Imputed Treatment
Employment Flag	0.33	0.01
Hours Worked Flag	0.36	0.04
Class Worker Flag	0.33	0.04
School Flag	0.28	0.01
Total Income Flag	0.42	0.05
N	87,315	531,160

Table 4.7: The Effects of DACA on Item- Nonresponse to Treatment Assignment Questions

Variables	(1) Education Flag	(2) Age Flag	(3) Year Immigrated Flag	(4) Citizenship Flag	(5) At Least One Flag
Eligible*Post	0.000	0.000	0.004	-0.007***	0.005
	(0.003)	(0.001)	(0.004)	(0.002)	(0.005)
Pre-DACA Mean	0.080	0.017	0.120	0.039	0.154
N	618,475	618,475	618,475	618,475	618,475
R^2	0.030	0.006	0.033	0.015	0.037

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.8: The Effects of DACA on Item- Nonresponse (Outcome of Interest)

Variables	(1) Employment Flag	(2) Hours Worked Flag	(3) Class Worker Flag	(4) School Flag	(5) Total Income Flag	(6) CATI or CAPI Interview
Whole Sample	-0.001	-0.004	0.006**	-0.003*	0.005*	0.015***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)
Pre-DACA Mean	0.047	0.078	0.070	0.038	0.087	0.716
Respondent Sample	-0.001	0.001	0.011***	0.001**	0.006***	0.015***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.005)
Pre-DACA Mean	0.013	0.036	0.034	0.007	0.041	0.709

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4.9: The Effects of DACA-Eligibility: Bounds Procedure

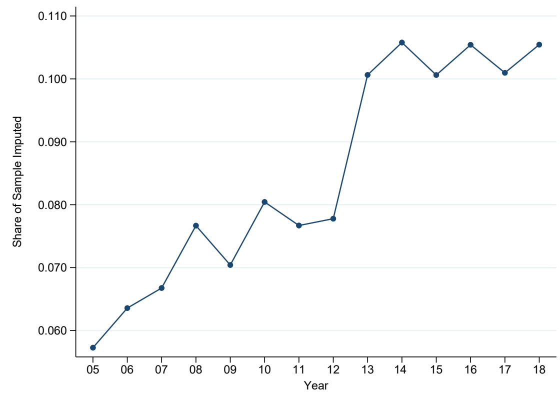
Variables	(1) Whole Sample	(2) Imputed as Eligible	(3) Imputed as Ineligible
Labor Force	0.019***	0.023***	0.027***
	(0.005)	(0.004)	(0.005)
Working	0.030***	0.032***	0.039***
	(0.005)	(0.004)	(0.005)
Unemployed	-0.016***	-0.013***	-0.019***
	(0.003)	(0.002)	(0.004)
Hours Worked	0.672**	0.857***	1.087***
	(0.261)	(0.218)	(0.256)
Self-Employed	-0.004	-0.002	-0.002
	(0.002)	(0.002)	(0.003)
School	0.005	0.011**	-0.005
	(0.005)	(0.005)	(0.005)
Total Income (IHS)	0.127**	0.187***	0.221***
	(0.055)	(0.047)	(0.055)

Standard errors in parenthesis

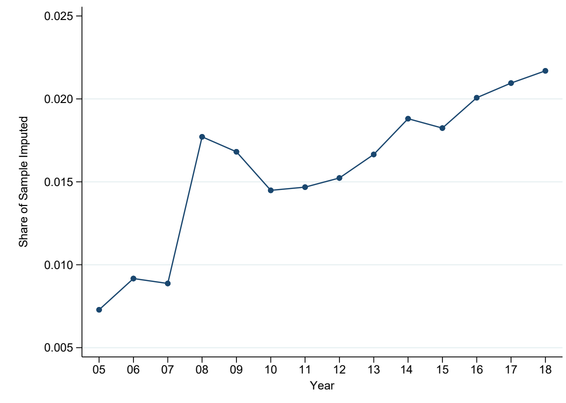
* $p < .10$, ** $p < .05$, *** $p < .01$

4.9 Chapter 4 Figures

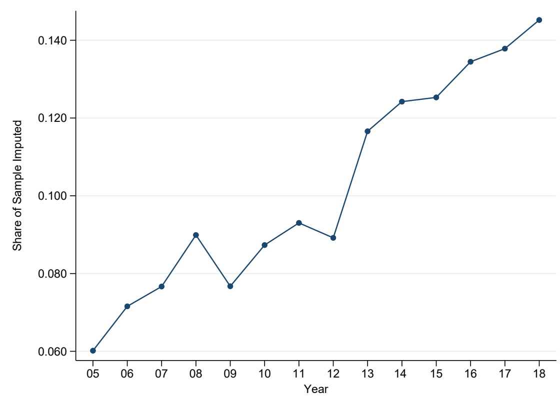
Figure 4.1: Share of Item Non-response to Questions used in Assigning Eligibility



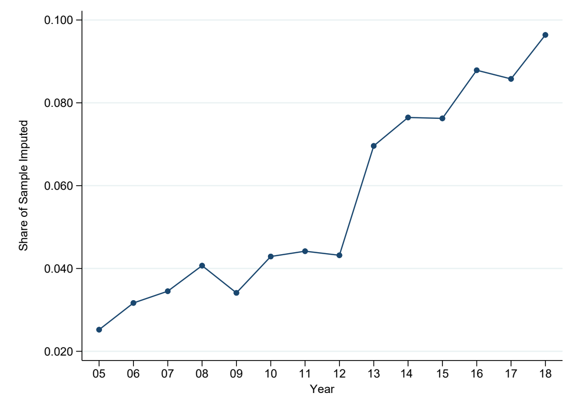
(A) Education Question



(B) Age Question



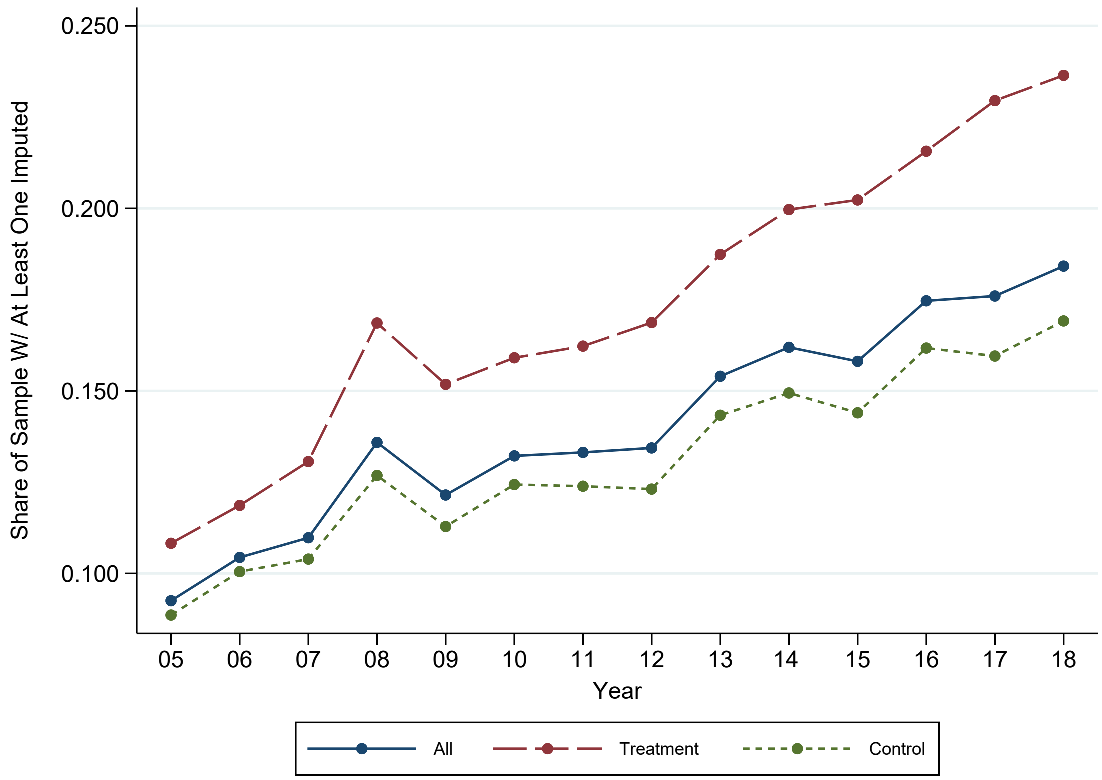
(C) Year Immigrated Question



(D) Citizen Question

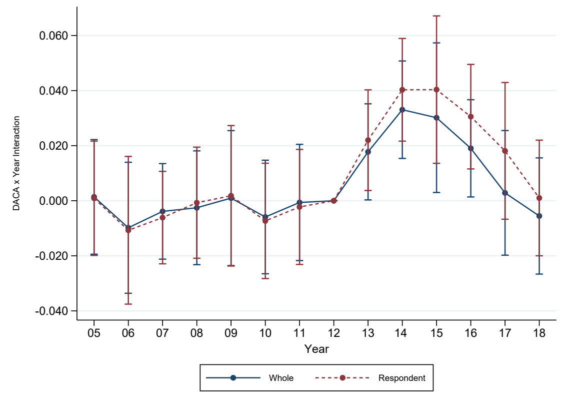
Notes: The figure plots the share of item non-response for each variable used for treatment assignment over time. The estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Source: Authors calculation from ACS.

Figure 4.2: Share of Treatment Imputed by Assignment Status

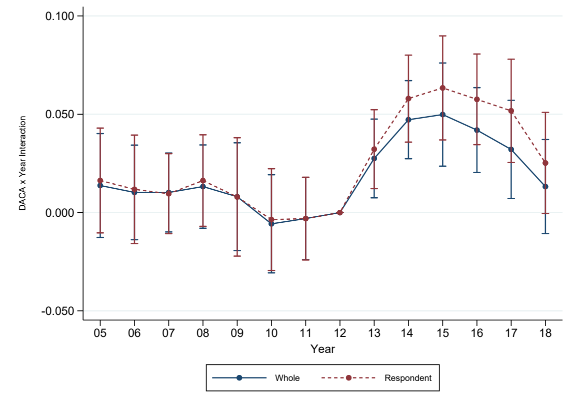


Notes: The figure plots the share of immigrants that had DACA-eligibility imputed by treatment status over time. The solid line shows the overall share of immigrants in the sample that had treatment imputed. The long dash line shows the share of immigrants assigned as DACA-eligible that had at least one of the key variables imputed. The short dash line shows the share of immigrants assigned as DACA-ineligible that had at least one of the key variables imputed. The estimates are derived from a sample of non-citizens ages 18-35 with at least a high school diploma (or equivalent). Source: Authors calculation from ACS.

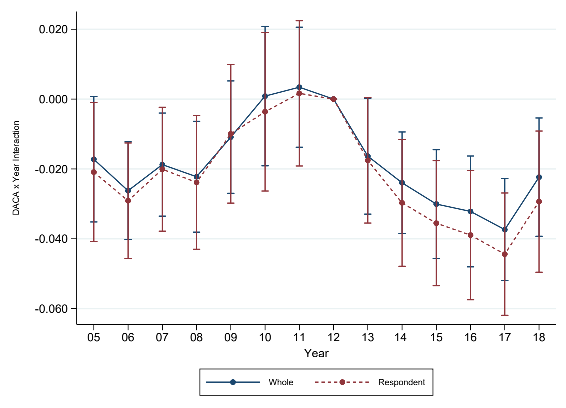
Figure 4.3: DACA and Labor Market Outcomes I: Respondent Sample vs. Whole Sample



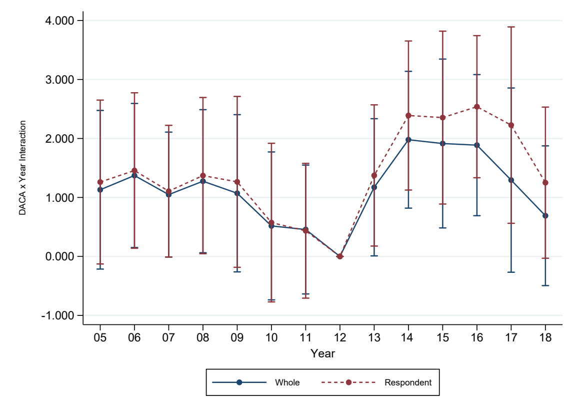
(A) In Labor Force



(B) Working



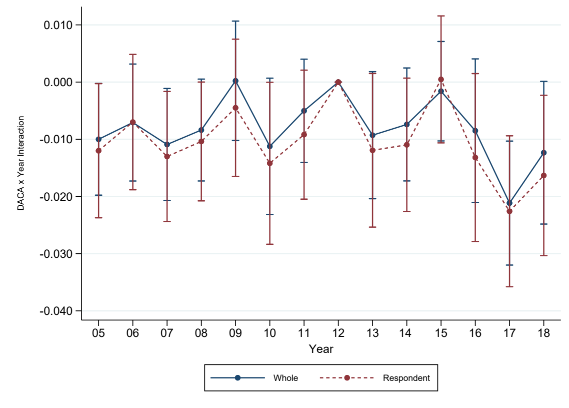
(C) Unemployed



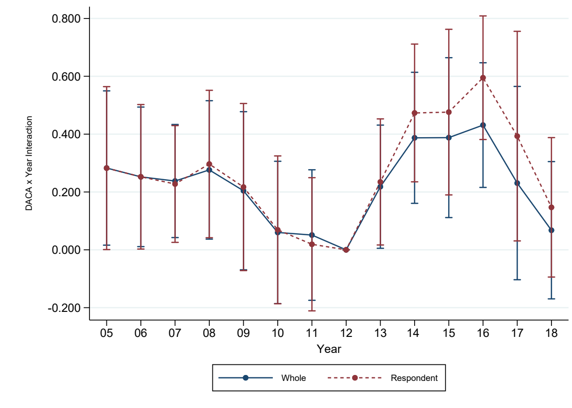
(D) Usual Hours Worked

Notes: The figure plots the coefficients obtained estimating Eq.(2) with the variable Eligible interacted with a binary variable for each year (2012 is the omitted interaction). 95% confidence limits of the interaction estimates are included in the graphs. Following dependent variables were used in the regressions (left-to-right, starting with the uppermost row): *In Labor Force* - binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months. The solid line estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent) who did not have their treatment status imputed. The dash line estimates are derived from a sample of all non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. Regressions control for DACA eligibility dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, Driver's license laws) and its interaction with the variable Eligible. Standard errors are clustered at state-year level.

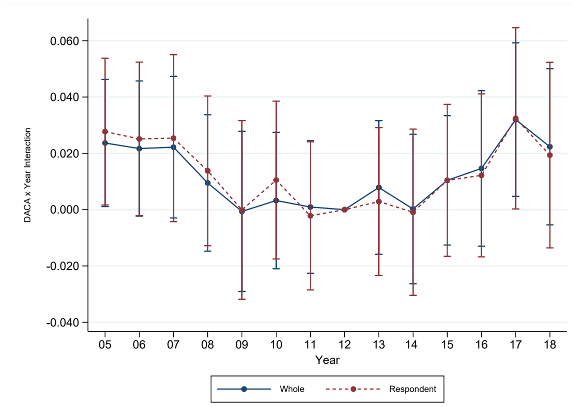
Figure 4.4: DACA and Labor Market Outcomes II: Respondent Sample vs. Whole Sample



(C) Self-Employed



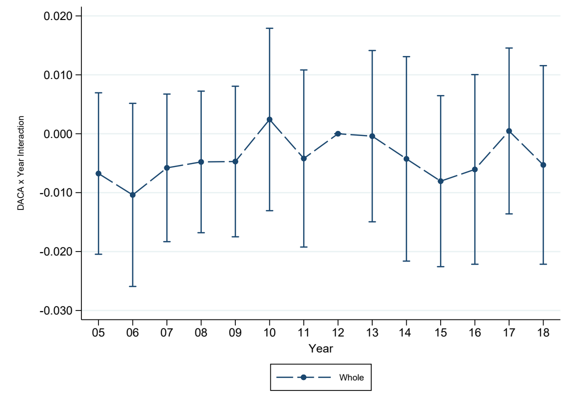
(D) total Income (IHS)



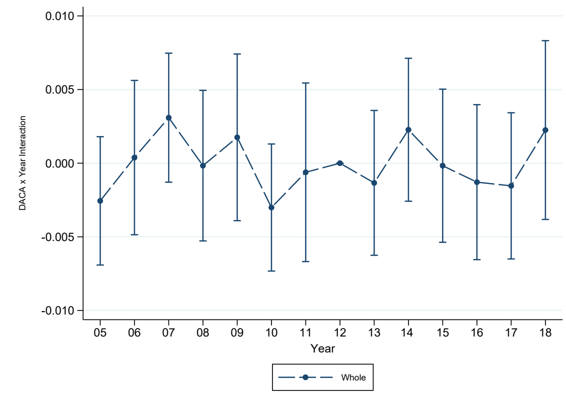
(E) In School

Notes: The figure plots the coefficients obtained estimating Eq.(2) with the variable Eligible interacted with a binary variable for each year (2012 is the omitted interaction). 95% confidence limits of the interaction estimates are included in the graphs. Following dependent variables were used in the regressions (left-to-right, starting with the uppermost row): *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *Income (IHS Transformation)* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months.; *School* - binary var. equal 1 if individual is currently attending school. The solid line estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent) who did not have their treatment status imputed. The dash line estimates are derived from a sample of all non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. Regressions control for DACA eligibility dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, Driver's license laws) and its interaction with the variable Eligible. Standard errors are clustered at state-year level.

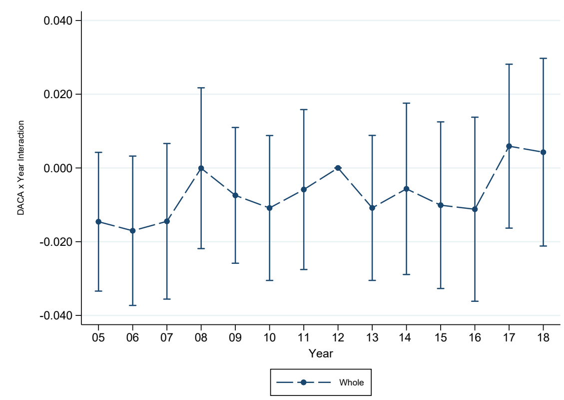
Figure 4.5: DACA on Survey-Item Response Rates in Assigning Eligibility Questions



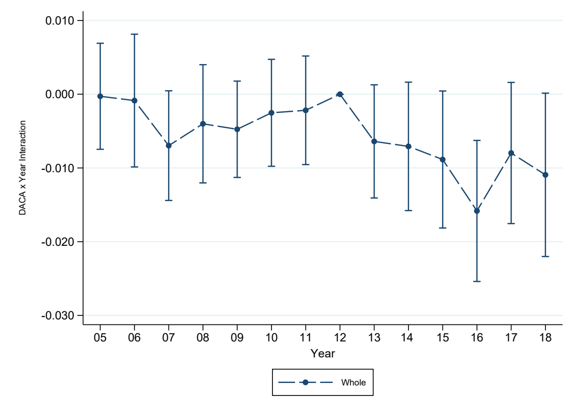
(A) Education Question



(B) Age Question



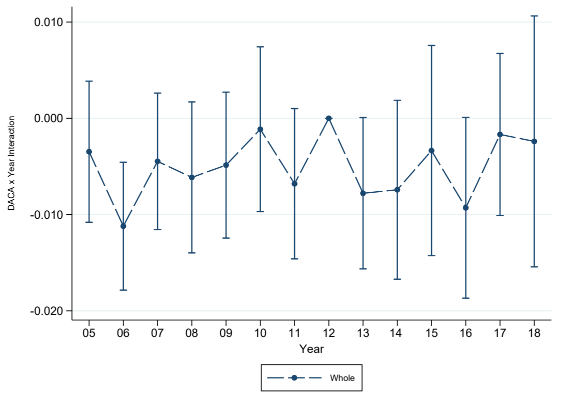
(C) Year Immigrated Question



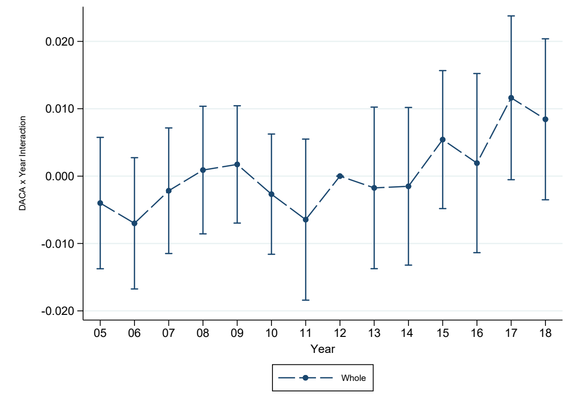
(D) Citizen Question

Notes: The figure plots the coefficients obtained estimating Eq.(2) with the variable Eligible interacted with a binary variable for each year (2012 is the omitted interaction). 95% confidence limits of the interaction estimates are included in the graphs. Following dependent variables were used in the regressions (left-to-right, starting with the uppermost row): *Education Question* - binary var. equal 1 if individual did not respond to the ACS question on Education; *Age Question* - binary var. equal 1 if individual did not respond to the ACS question on Age; *Year Immigrated Question* - binary var. equal 1 if individual did not respond to the ACS question on Year Immigrated; *Citizenship Question* - binary var. equal 1 if individual did not respond to the ACS question on Citizenship. All estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. Regressions control for DACA eligibility dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, Driver's license laws) and its interaction with the variable Eligible. Standard errors are clustered at state-year level.

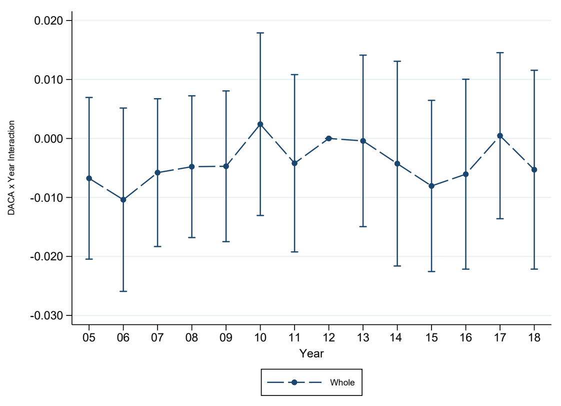
Figure 4.6: DACA on Survey-Item Response Rates in Outcome Questions



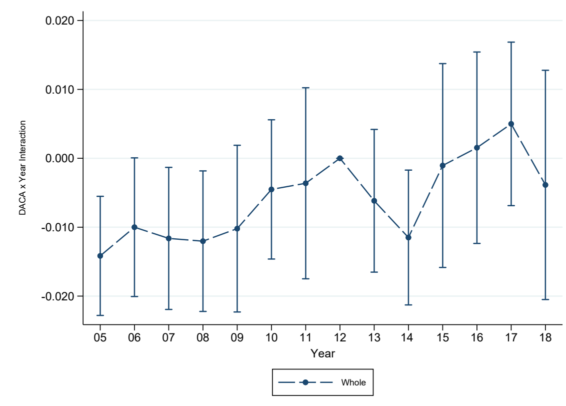
(A) Employment Question



(B) Class of Worker Question



(C) School Question



(D) Total Income Question

Notes: The figure plots the coefficients obtained estimating Eq.(2) with the variable Eligible interacted with a binary variable for each year (2012 is the omitted interaction). 95% confidence limits of the interaction estimates are included in the graphs. Following dependent variables were used in the regressions (left-to-right, starting with the uppermost row): *Employment Question* - binary var. equal 1 if individual did not respond to the ACS question on Employment; *Class of Worker Question* - binary var. equal 1 if individual did not respond to the ACS question on Class of Worker; *School Question* - binary var. equal 1 if individual did not respond to the ACS question on school attendance; *Total Income Question* - binary var. equal 1 if individual did not respond to the ACS question on total income. All estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. Regressions control for DACA eligibility dummy, demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age, age of entering U.S., education attainment), state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, state minimum wage, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, Driver's license laws) and its interaction with the variable Eligible. Standard errors are clustered at state-year level.

Appendices

Appendix A: Chapter 1 Supplementary Tables

Table A.1: Descriptive Statistics of the Constructed Probability by State-of-residence

State of Residence	(1) Eligible Population	(2) Total Approvals	(3) Constructed Probability
California	467,454	229,335	0.491
Texas	278,271	127,555	0.458
Florida	135,568	35,682	0.263
New York	129,052	45,335	0.351
Illinois	83,039	43,443	0.523
New Jersey	61,906	23,471	0.379
Arizona	57,769	28,483	0.493
Georgia	51,906	24,907	0.480
North Carolina	43,668	27,866	0.638
Washington	43,256	18,483	0.427
Virginia	34,642	12,821	0.370
Colorado	32,330	17,676	0.547
Maryland	29,534	10,367	0.351
Massachusetts	28,637	8,687	0.303
Nevada	28,524	13,379	0.469
Pennsylvania	25,540	6,500	0.254
Oregon	21,066	11,534	0.548
Michigan	19,364	6,840	0.353
Utah	18,102	9,878	0.546
Connecticut	17,416	5,265	0.302

Notes: The DACA-eligible population for each state-of-residence is estimated using the eligiblity procedure defined in Section 4. Total DACA approvals by state-of-residence up to march 2018 come from the publicly available USCIS data. The constructed probability is constructed by the author by dividing total approvals by estimated size of the observed DACA-eligible population for each state-of-residence. The 20 states with largest eligible population are displayed.

Table A.2: The Effects of DACA on Labor Market Outcomes

State-of-Residence Variation

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(6) School
DACA Recipient	0.085***	0.110***	-0.038***	3.109***	-0.007	0.009
	(0.013)	(0.013)	(0.007)	(0.670)	(0.006)	(0.013)
Pre-DACA Mean	0.747	0.662	0.113	27.929	0.050	0.251
N	618,450	618,450	432,284	618,450	498,081	618,450
R^2	0.151	0.146	0.034	0.201	0.029	0.333

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level

Table A.3: The Effects of DACA on Income (IHS)

State-of-Residence Variation

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
DACA Recipient	0.741***	0.871***	-0.152***	-0.019	-0.113***
	(0.145)	(0.139)	(0.041)	(0.015)	(0.028)
Pre-DACA Mean	$15,117	$14,165	$245	$40	$12.63
N	618,450	618,450	618,450	618,450	423,477
R^2	0.181	0.158	0.026	0.014	0.300

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. *Wage* - inverse hyperbolic sine (IHS) transformation of constructed wage. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table A.4: Placebo Test on Naturalized Citizens (Labor Market)

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(6) School
$P(R\|C,T,E)$	0.019*	0.017	0.002	0.225	0.005	0.001
	(0.009)	(0.010)	(0.005)	(0.384)	(0.006)	(0.010)
$P(R\|S,T,E)$	-0.001	0.001	-0.002	-0.635	-0.004	-0.002
	(0.009)	(0.010)	(0.005)	(0.393)	(0.006)	(0.008)

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of naturalized immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level

Table A.5: Placebo Test on Naturalized Citizens (Income)

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
$P(R\|C,T,E)$	0.165*	0.087	-0.016	-0.005	-0.073
	(0.082)	(0.101)	(0.041)	(0.016)	(0.085)
$P(R\|S,T,E)$	-0.093	-0.112	-0.085*	-0.014	-0.220*
	(0.093)	(0.108)	(0.038)	(0.017)	(0.101)

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. *Wage* - inverse hyperbolic sine (IHS) transformation of constructed wage. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Table A.6: The Effects of DACA on Labor Market Outcomes

Lubotsky and Wittenberg (2006) Approach

Variables	(1) Labor Force	(2) Working	(3) Unemployed	(4) Usual Hours Worked	(5) Self-Employed	(6) School
$\rho = \hat{\rho}$	0.095***	0.113***	-0.025 ***	4.005 ***	-0.007	0.223***
	(0.012)	(0.012)	(0.006)	(0.570)	(0.006)	(0.042)
$\hat{\rho}$	0.774	0.754	0.481	0.532	1.007	- 2.265
$\rho = 1$	0.090***	0.115***	-0.039 ***	3.329***	-0.007	0.012
	(0.013)	(0.014)	(0.007)	(0.698)	(0.006)	(0.013)

Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

Notes: *Labor Force* -binary var. equal 1 if individual currently participates in the labor force; *Working* - binary var. equal 1 if individual is currently working; *Unemployed* - binary var. equal 1 if individual is currently unemployed; *Usual Hours Worked* - number of hours per week respondent usually worked in past 12 months; *Self-Employed* - binary var. equal 1 if individual is currently self-employed. *School* - binary var. equal 1 if individual is currently in school. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control for demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level

Table A.7: The Effects of DACA on Income (IHS)

Lubotsky and Wittenberg (2006) Approach

Variables	(1) Total Income	(2) Wage Income	(3) Other Income	(4) Welfare Income	(5) Wage
$\rho = \hat{\rho}$	0.960***	1.023 ***	-0.029	-0.020	-0.042
	(0.136)	(0.136)	(0.035)	(0.016)	(0.027)
$\hat{\rho}$	0.772	0.801	0.711	1.065	0.828
$\rho = 1$	0.809***	0.9344 ***	-0.140***	-0.020	-0.104 ***
	(0.152)	(0.146)	(0.041)	(0.015)	(0.029)

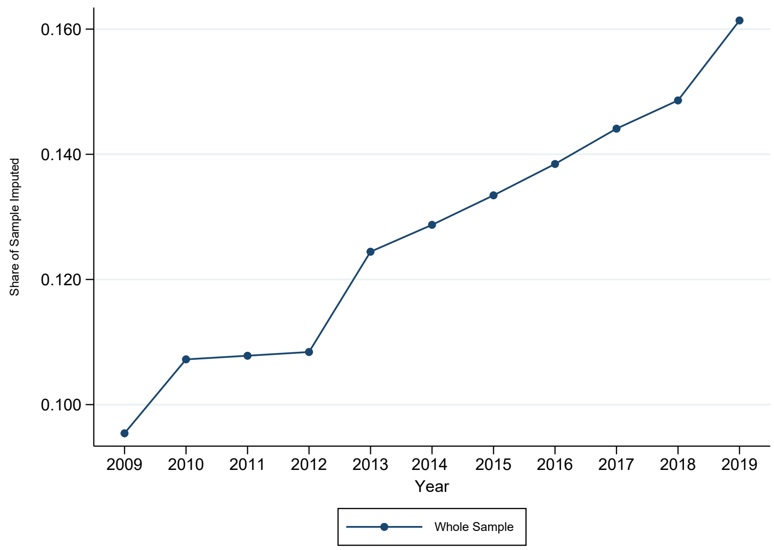
Standard errors in parenthesis

* $p < .10$, ** $p < .05$, *** $p < .01$

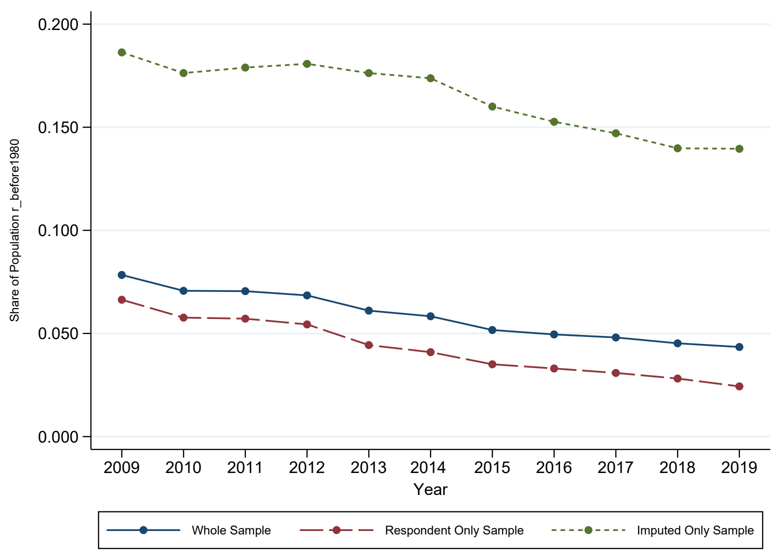
Notes: *Total Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax personal income from all sources for the previous 12 months. *Wage Income* - inverse hyperbolic sine (IHS) transformation of individual's total pre-tax wage and salary income for the previous 12 months. *Other Income* - inverse hyperbolic sine (IHS) transformation of individual's total income from sources not included in the other IPUMS person-record income variables for the previous 12 months. *Welfare Income* - inverse hyperbolic sine (IHS) transformation of individual's pre-tax income (if any) the respondent received from various public assistance programs commonly referred to as "welfare" during the previous 12 months. *Wage* - inverse hyperbolic sine (IHS) transformation of constructed wage. Estimates are derived from a sample of non-citizen immigrants ages 18-35 with at least a high school diploma (or equivalent). Data are taken from the 2005-2018 waves of ACS. All regressions control demographic characteristics (sex, race, ethnicity, marital status), DACA eligibility criteria dummies (age,age of entering U.S., education attainment), country-of-origin fixed effects, state fixed effects, year fixed effects, state-specific time trends, and state-year controls (state unemployment, dummies for Universal E-Verify laws, Public E-Verify, instate tuition laws, driver's license laws). Standard errors are clustered at state-year level.

Appendix B: Chapter 2 Supplementary Figures

Figure B.1: Year Immigrated Condition



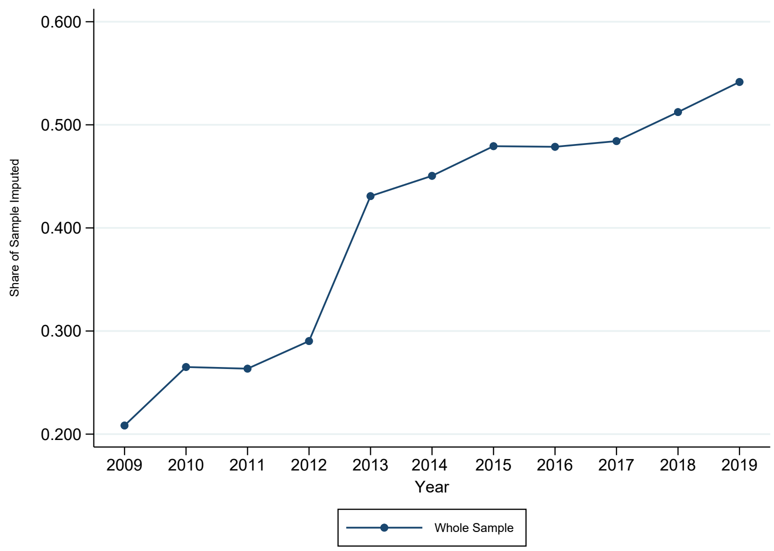
(A) Share of Sample with Imputed Year Immigrated Question



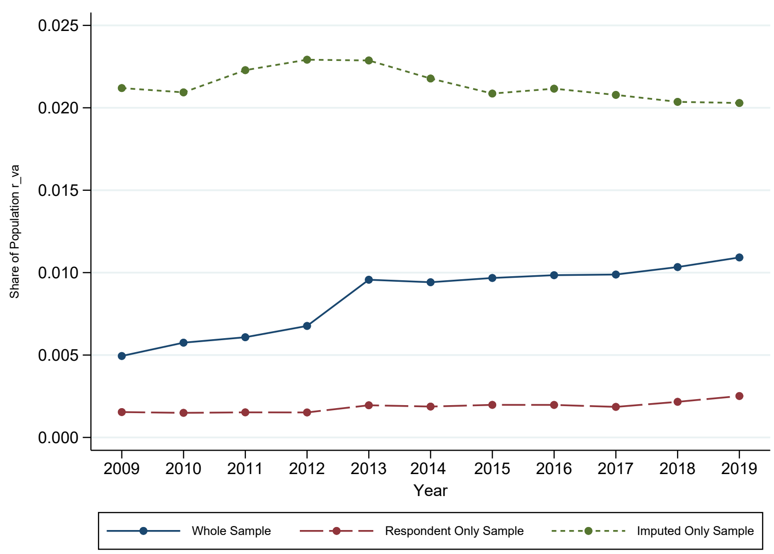
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.2: Veteran's Insurance Condition



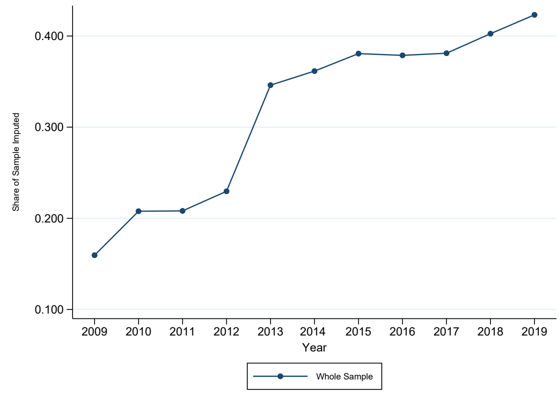
(A) Share of Sample with Imputed Veteran's Insurance Question



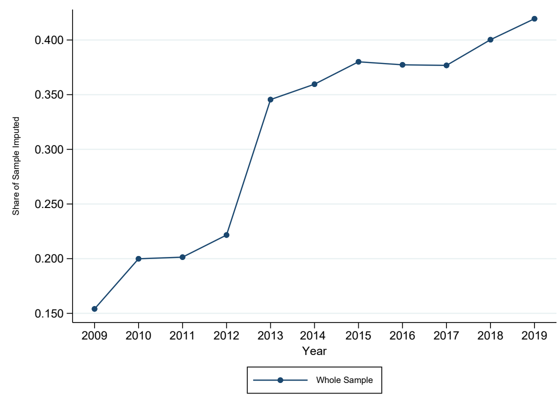
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.3: Active Military or Veteran Condition



(A.1.) Share of Sample with Imputed Employment Status Question

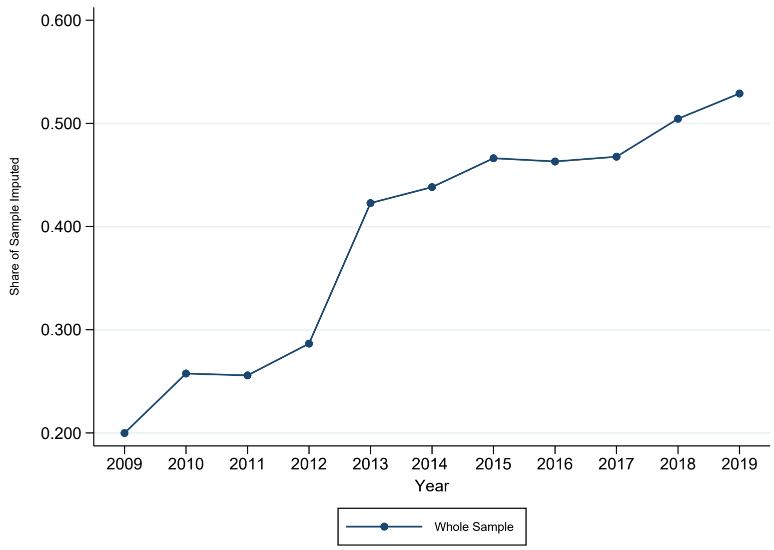


(A.2.) Share of Sample with Imputed Veteran Status Question

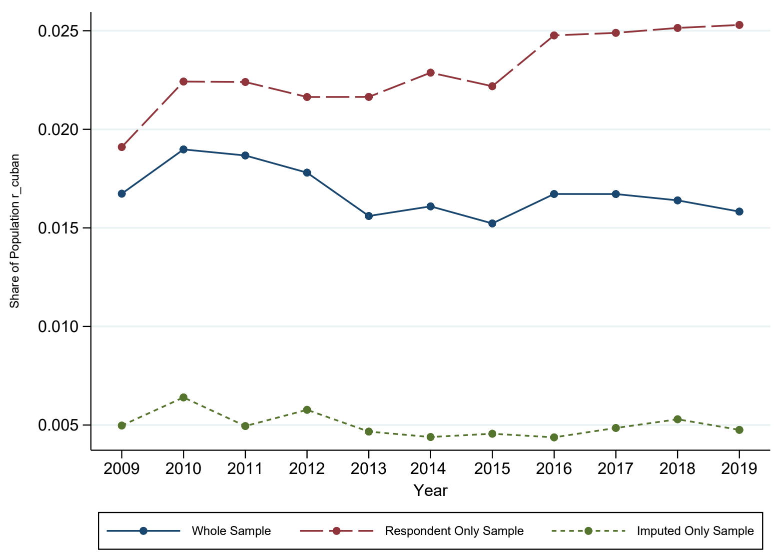
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.4: Born in Cuba Condition



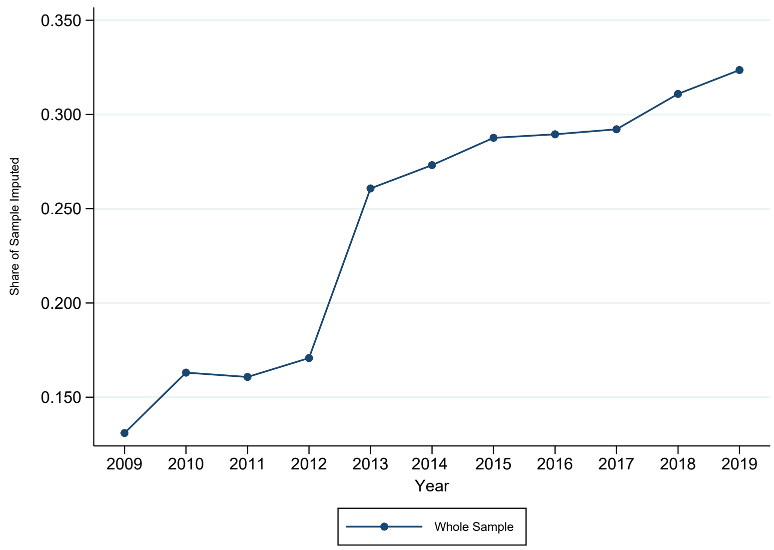
(A) Share of Sample with Imputed Place-of-Birth Question



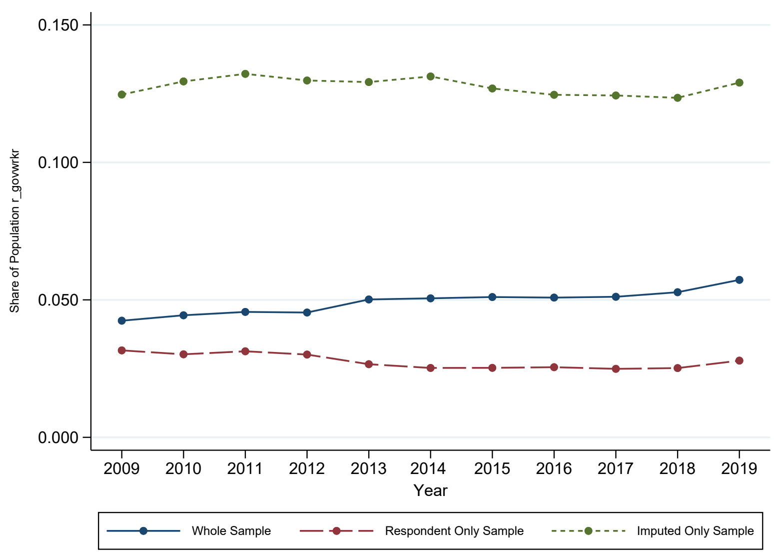
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.5: Government Employee Condition



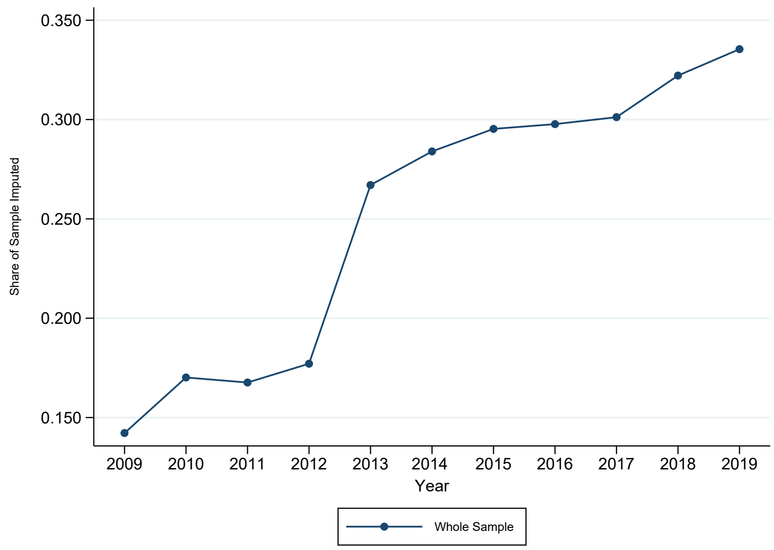
(A) Share of Sample with Imputed Class of Worker Question



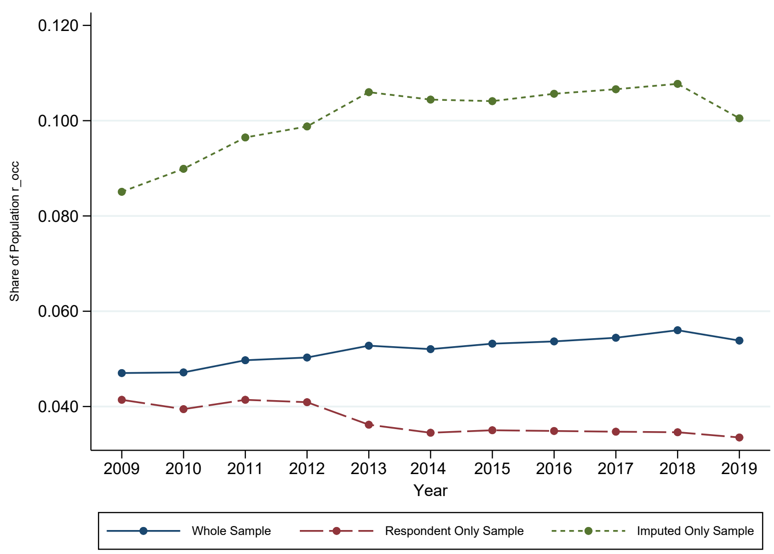
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.6: Type of Occupation Condition



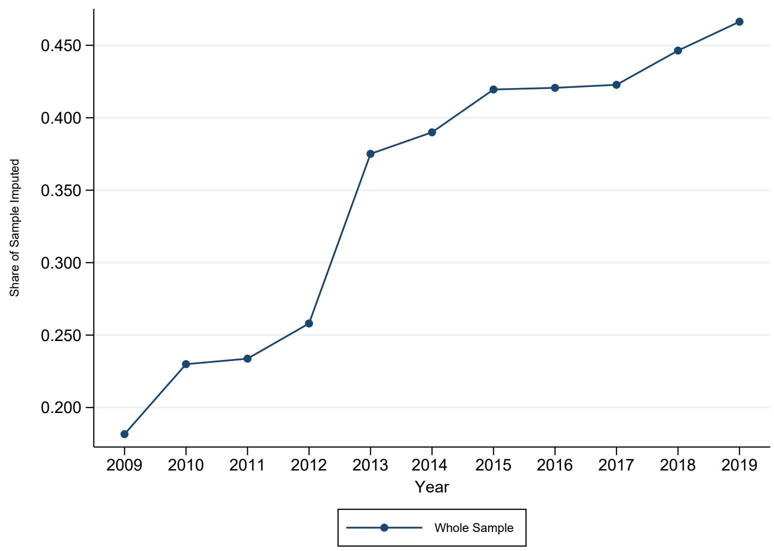
(A) Share of Sample with Imputed Occupation Question



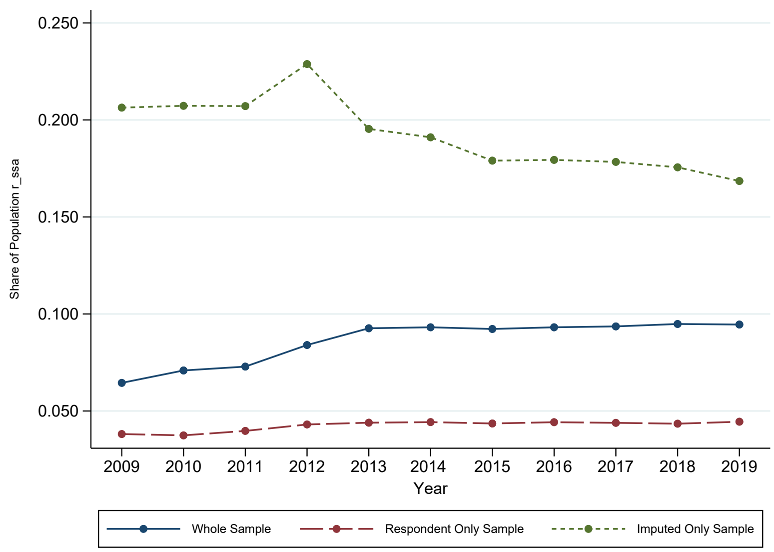
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.7: Social Security Income Condition



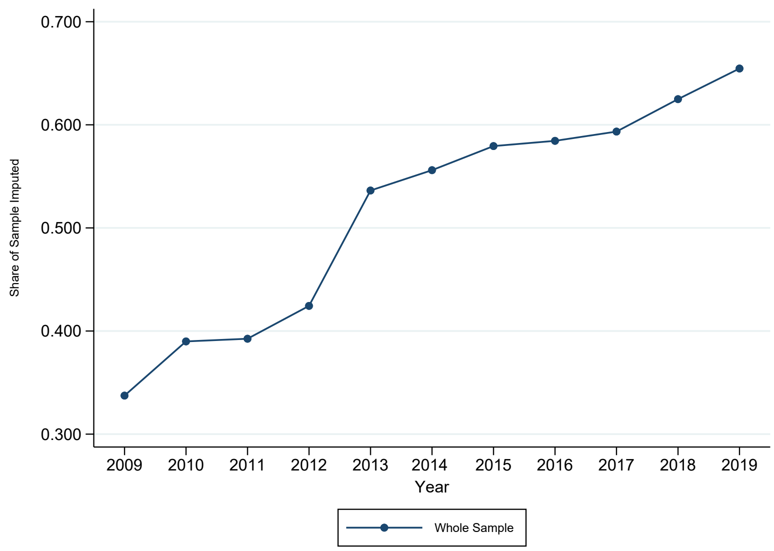
(A) Share of Sample with Imputed Social Security Income Question



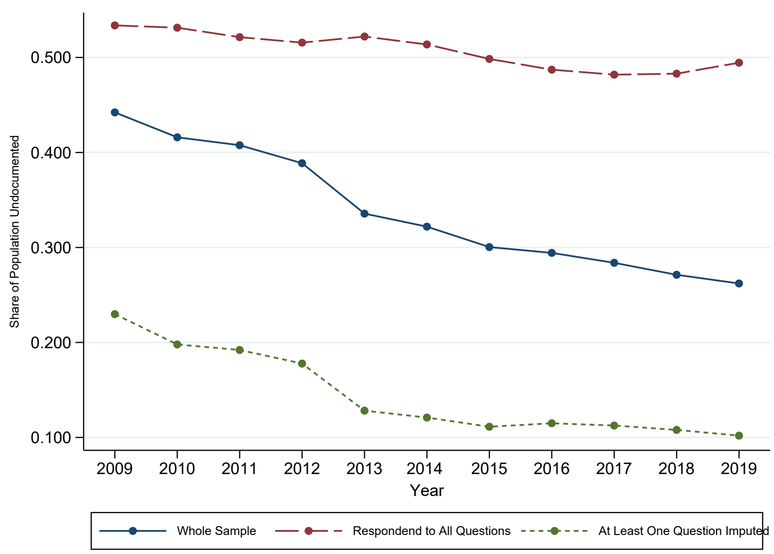
(B) Share of Sample that Satisfy condition by Question Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by question response status are calculated using census person weights.

Figure B.8: All Conditions



(A) Share of Sample with at least one imputed Question



(B) Share of Sample Undocumented by Questions Response Status

Notes: Author's own calculations using the American Community Survey. Sample is composed of all individuals that responded to the citizenship question as non-citizens and that those that did not respond to the citizenship question. Share of sample imputed is unweighted. Share of population by response status are calculated using census person weights.

Appendix C: Chapter 2 ACS Imputation Procedure for Citizenship Question

This section details the imputation procedure as provided by the Census to IPUMS. This imputation procedure is likely incomplete. The Census uses geographic information for the hot-deck imputation procedure of demographic information which is not mentioned here. As the Census does not provide a publicly available document detailing every step of the imputation procedure, it is not possible to know what or if any additional steps are taken by the Census in imputing item nonresponse in the citizenship question.

The Census imputation procedure provided to and released by IPUMS for citizenship, year of immigration, and year naturalized in the ACS is as follows:

- If a person reports being born in the United Sates when asked their birthplace (BPL) but reports not being born in the U.S. when asked if they are a U.S. citizen (CITIZEN), CITIZEN will be replaced with"Born in the U.S." When this happens, QCITIZEN will show the value is allocated.

- If year of immigration (YRIMMIG) is one year after the survey year, YRIMMIG will be replaced with the survey year.

- If a person reports being born in Puerto Rico, Guam, Northern Marianas, or the Virgin Islands when asked their birthplace (BPL) and either does not have a response for when asked about their citizenship, says they are a citizen but does not specify what type, says they were born in the U.S., or says they are not a citizen, CITIZEN will be replaced with "Born in the Puerto Rico, etc." When this happens, QCITIZEN will show the value is allocated.

- If a person is foreign-born (BPL) and either does not have a response for when asked about their citizenship, says they are a citizen but does not specify what type, says they are a citizen who was born in the U.S., or says they are a citizen who was born in Puerto Rico (CITIZEN), CITIZEN will be allocated based on their parents citizenship. If the person has a parent in the household who is US-born, CITIZEN will be replaced with "Born abroad of American parents." If the parent is a naturalized citizen, CITIZEN will be replaced with "Naturalized citizen." If the parent is not a citizen, CITIZEN will be replaced with "Not a citizen." When this happens, QCITIZEN will show the value is allocated.

  - RELATE is used to determine parents: A person with value of "parent" in RELATE is the parent to the reference person or brother/sister. The reference person and spouse are the parents to the son/daughter or foster child. The son/daughter or foster child of the reference person are parents to grandchildren of the reference person.

- If after the previous edits, a person still has a value of "Yes" for being a citizen, but does not specify which type of citizen or is missing (CITIZEN), and year of immigration is equal to or after the year they were born, CITIZEN will be allocated - the allocated value will be drawn from another person with the same age, race, and ethnicity. If year of immigration is also missing or is prior to when a person was born, CITIZEN and YRIMMIG will be allocated jointly - these values will be drawn from another person with the same age, race, and ethnicity. When this happens, QCITIZEN and/or QYRIMM will show the values are allocated.

- If after the previous edits, a person indicates they are a citizen who was born in the U.S. or Puerto Rico but lists a foreign birthplace (BPL), and year of immigration is equal to or after the year they were born, CITIZEN will be allocated - the allocated value is drawn from another person of a similar age, race, and ethnicity. If year of immigration is also missing or is prior to when a person was born, CITIZEN and YRIMMIG will be allocated jointly from another person with a similar age, race, and ethnicity. When this happens, QCITIZEN and/or QYRIMM will show the values are allocated.

- If a person reports being born in Puerto Rico, Guam, Northern Marianas, or the Virgin Islands when asked their birthplace (BPL) and says they are a citizen who was born abroad to American parents or says they are a naturalized citizen, CITIZEN will be replaced with "Born in the Puerto Rico, etc." When this happens, QCITIZEN will show the value is allocated.

- If after the previous edits, a foreign-born (BPL) person still indicates being a naturalized citizen (CITIZEN), and year of immigration is after they[the] year they were born and either the same year as the survey year or the year before the survey year, CITIZEN will be replaced with "Not a citizen." When this happens, QCITIZEN will show the value is allocated.

- If a person reports being a citizen and born in the U.S. and they current live in one of the 50 states (STATE) or they report being a citizen who was born in Puerto Rico and they currently live in Puerto Rico, YRIMMIG will be replaced with "Not in universe."

- For respondents who are in universe for having a year of immigration, if YRIMMIG is not reported or if it is before the year a person was born, YRIMMIG will be allocated from another person with a similar age, race, and ethnicity. When this happens, QYRIMM will show the value is allocated.

- Beginning in 2008, if a person reports their year of naturalization as prior to 1883, reports being a naturalized citizen but leaves the year blank, reports a year of naturalization before they were born, or reports a year of naturalization after the survey year, YRNATUR will be allocated from someone else with a similar age (AGE), race (RACE), and ethnicity (HISPAN).

- Beginning in 2008, if a person reports not being a U.S. citizen or being born in the U.S., U.S. territories, or abroad to U.S. parents (CITIZEN) and reports a value for year of naturalization, YRNATUR will be replaced with a missing value.

Bibliography

Aigner, D. J. (1973). Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics*, 1(1):49–59.

Amuedo-Dorantes, C. and Antman, F. (2017). Schooling and labor market effects of temporary authorization: Evidence from daca. *Journal of population economics*, 30(1):339–373.

Amuedo-Dorantes, C., Bansak, C., and Raphael, S. (2007). Gender differences in the labor market: Impact of irca. *American Economic Review*, 97(2):412–416.

Baker, B. (2021). Estimates of the unauthorized immigrant population residing in the united states: January 2015–january 2018. *Population Estimates*.

Baker, B. and Rytina, N. (2013). Estimates of the unauthorized immigrant population residing in the united states: January 2012. *Department of Homeland Security, March*. Link.

Baker, S. R. (2015). Effects of immigrant legalization on crime. *American Economic Review*, 105(5):210–13.

Ballis, B. et al. (2020). Does peer motivation impact educational investments? evidence from daca.

Batalova, J., Hooker, S., and Capps, R. (2014). Daca at the two-year mark: A national and state profile of youth eligible and applying for deferred action. *Migration Policy Institute*.

Bollinger, C. R. and Hagstrom, P. (2008). Food stamp program participation of refugees and immigrants. *Southern Economic Journal*, pages 665–692.

Bollinger, C. R. and Hagstrom, P. (2011). The poverty reduction success of public transfers for working age immigrants and refugees in the united states. *Contemporary Economic Policy*, 29(2):191–206.

Bollinger, C. R. and Hirsch, B. T. (2006). Match bias from earnings imputation in the current population survey: The case of imperfect matching. *Journal of Labor Economics*, 24(3):483–519.

Borjas, G. J. (2017). The labor supply of undocumented immigrants. *Labour Economics*, 46:1–13.

Borjas, G. J. and Cassidy, H. (2019). The wage penalty to undocumented immigration. *Labour Economics*, 61:101757.

Brown, J. D., Heggeness, M. L., Dorinski, S. M., Warren, L., Yi, M., et al. (2018). *Understanding the quality of alternative citizenship data sources for the 2020 census.* US Census Bureau, Center for Economic Studies Washington, DC.

Bureau, U. C. (2014). American community survey design and methodology.

Capps, R., Gelatt, J., Van Hook, J., and Fix, M. (2018). Commentary on "the number of undocumented immigrants in the united states: Estimates based on demographic modeling with data from 1990-2016". *PloS one*, 13(9):e0204199.

Clark, S. (2014). 2014 american community survey research and evaluation report memorandum series acs14-rer-26.

Fazel-Zarandi, M. M., Feinstein, J. S., and Kaplan, E. H. (2018). The number of undocumented immigrants in the united states: Estimates based on demographic modeling with data from 1990 to 2016. *PloS one*, 13(9):e0201193.

Giuntella, O. and Lonsky, J. (2018). The effects of daca on health insurance, access to care, and health outcomes. *Available at SSRN 3170258*. Link.

Giuntella, O. and Lonsky, J. (2020). The effects of daca on health insurance, access to care, and health outcomes. *Journal of Health Economics*, page 102320.

Giuntella, O., Lonsky, J., Mazzonna, F., and Stella, L. (2021). Immigration policy and immigrants' sleep. evidence from daca. *Journal of Economic Behavior & Organization*, 182:1–12.

Gonzalez, R. G. and Bautista-Chavez, A. M. (2014). *Two Years and Counting: Assessing the Power of DACA.* American Immigration Council. Link.

Grassley, C. (2017). Data indicate unauthorized immigrants exploited loophole to gain legal status, pathway to citizenship.

Grigorieff, A., Roth, C., and Ubfal, D. (2020). Does information change attitudes toward immigrants? *Demography*, 57(3):1117–1143.

Gunadi, C. (2019). Does immigrant legalization affect crime? evidence from deferred action for childhood arrivals in the united states. Link.

Hainmueller, J., Lawrence, D., Martén, L., Black, B., Figueroa, L., Hotard, M., Jiménez, T. R., Mendoza, F., Rodriguez, M. I., Swartz, J. J., et al. (2017). Protecting unauthorized immigrant mothers improves their children's mental health. *Science*, 357(6355):1041–1044.

Hipsman, F., Gómez-Aguiaga, B., and Capps, R. (2016). Daca at four: Participation in the deferred action program and impacts on recipients.". *Notes*, 1(398,000):228–000. Link.

Hirsch, B. T. and Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. *Journal of labor economics*, 22(3):689–722.

Hsin, A. and Ortega, F. (2018). The effects of deferred action for childhood arrivals on the educational outcomes of undocumented students. *Demography*, 55(4):1487–1506.

Kossoudji, S. A. and Cobb-Clark, D. A. (2000). Irca's impact on the occupational concentration and mobility of newly-legalized mexican men. *Journal of Population Economics*, 13(1):81–98.

Kuka, E., Shenhav, N., and Shih, K. (2019). A reason to wait: The effect of legal status on teen pregnancy. In *AEA Papers and Proceedings*, volume 109, pages 213–17.

Kuka, E., Shenhav, N., and Shih, K. (2020). Do human capital decisions respond to the returns to education? evidence from daca. *American Economic Journal: Economic Policy*, 12(1):293–324.

Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica*, 75(2):537–551.

Lozano, F. A. and Sorensen, T. (2011). The labor market value to legal status.

Lubotsky, D. and Wittenberg, M. (2006). Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics*, 88(3):549–562.

Manski, C. F. (2011). Policy analysis with incredible certitude. *The Economic Journal*, 121(554):F261–F289.

Manski, C. F. (2016). Credible interval estimates for official statistics with survey nonresponse. *Journal of Econometrics*, 191(2):293–301.

Nguimkeu, P., Denteh, A., and Tchernis, R. (2019). On the estimation of treatment effects with endogenous misreporting. *Journal of econometrics*, 208(2):487–506.

O'Hare, W. P. (2018). Citizenship question nonresponse.

Passel, J. S. and Cohn, D. (2019). Mexicans decline to less than half the us unauthorized immigrant population for the first time. *Pew Research Center*. Link.

Passel, J. S. and Lopez, M. H. (2012). Up to 1.7 million unauthorized immigrant youth may benefit from new deportation rules. *Washington, DC: Pew Hispanic Center*. Link.

Pope, N. G. (2016). The effects of dacamentation: The impact of deferred action for childhood arrivals on unauthorized immigrants. *Journal of Public Economics*, 143:98–114.

Ro, A. and Van Hook, J. (2021). Comparing the effectiveness of assignment strategies for estimating likely undocumented status in secondary data sources for latino and asian immigrants. *Population Research and Policy Review*, pages 1–16.

Rosenblum, M. R. (2015). Understanding the potential impact of executive action on immigration enforcement. *Migration Policy Institute*, page 11.

Ruggles, S., Flood, S., Goeken, R., Grover, J., and Meyer, E. (2020). Jose pacas, and matthew sobek. 2020. ipums usa: Version 10.0 [dataset]. *Minneapolis, MN: IPUMS*, 10:D010.

Rytina, N. (2002). *IRCA legalization effects: Lawful permanent residence and naturalization through 2001*. Office of Policy and Planning, Statistics Division, US Immigration and Naturalization Service. Link.

SSA (2018). Social security benefits paid to non-citizens granted deferred action for childhood arrivals status.

USCIS (2013-2017b). Deferred action for childhood arrivals (daca) quarterly report.

USCIS (2017a). Deferred action for childhood arrival (daca) recipients serving in the military.

USCIS (2018). The department of homeland security's response to representative king's january 16, 2018 letter on daca characteristics data.

USCIS (2019). Daca characteristics data: Approximate active daca recipients as of nov. 30, 2018.

Van Hook, J. and Bachmeier, J. D. (2013). How well does the american community survey count naturalized citizens? *Demographic research*, 29(1):1.

Van Hook, J., Bachmeier, J. D., Coffman, D. L., and Harel, O. (2015). Can we spin straw into gold? an evaluation of immigrant legal status imputation approaches. *Demography*, 52(1):329–354.

Van Hook, J., Morse, A., Capps, R., and Gelatt, J. (2021). Uncertainty about the size of the unauthorized foreign-born population in the united states. *Demography*.

Venkataramani, A. S., Shah, S. J., O'Brien, R., Kawachi, I., and Tsai, A. C. (2017). Health consequences of the us deferred action for childhood arrivals (daca) immigration programme: a quasi-experimental study. *The Lancet Public Health*, 2(4):e175–e181.

Warren, R. (2020). Reverse migration to mexico led to us undocumented population decline: 2010 to 2018. *Journal on Migration and Human Security*, 8(1):32–41.

Warren, R. and Passel, J. S. (1987). A count of the uncountable: estimates of undocumented aliens counted in the 1980 united states census. *Demography*, pages 375–393.

Vita

Andrés Felipe Mira

Place of Birth: Bogota, Colombia

Education:

North Carolina State University	Raleigh, NC
Masters in Economics	May. 2017
University of Central Florida	Orlando, FL
B.A. in Economics	May. 2015

Professional Positions:

Graduate Teaching Assistant, University of Kentucky	Fall 2021–Spring 2022
Research Assistant, University of Kentucky	Fall 2020–Spring 2021
Graduate Teaching Assistant, University of Kentucky	Fall 2017–Spring 2020

Honors

Poster Award, APPAM Fall Research Conference	2022
Mark Berger Best Paper Award, KEA Annual Conference	2021
Luckett Fellowship, University of Kentucky	2021–2022
Gatton Doctoral Fellowship, University of Kentucky	2020–2021
ISFE Independent Research Award, University of Kentucky	Summer 2020
Gatton Doctoral Fellowship, University of Kentucky	2018–2019
Lyman T. Johnson Fellowship, University of Kentucky	2017–2020
Max Steckler Fellowship, University of Kentucky	2017–2018