Online Appendices (Not for Publication)

Opening the Door: Immigrant Legalization and Family Reunification in the United States

Elizabeth U. Cascio
Ethan Lewis

Appendix A. Data

I. Treatment Variable: the Legalization Ratio

A. Legalization Applications Processing System (LAPS) data

The SAW and GLP admissions that enter the numerator of the legalization ratio were taken from the Legalization Applications Processing System (LAPS), available from the National Archives. These public-use microdata consist of selected fields from anonymized records from all forms I-687 (application for temporary legal status under IRCA's general legalization program, spilt across two files) and forms I-700 (application for temporary legal status under IRCA's SAW program) received by the Immigration and Naturalization Service (INS), consisting of 3,040,948 records in total.

These fields describe some outcomes of the application process, including whether and when a Green Card was awarded, through the end of the 1992 fiscal year.[1] This is critical to establishing the timing for our event-study model, as outlined in Figure 2. These fields also include the applicant's country of birth and state and county of intended residence within the U.S. (current U.S. address) at the time of application. In these and all other administrative data, we code counties to metropolitan areas using 1999 Primary Metropolitan Statistical Areas (PMSA) boundaries.[2] For the metro area-level analysis for Mexican admissions, we focus on 66 metropolitan areas that are observable in admissions statistics published by Department of Homeland Security (DHS) for years 2007 and later.[3] For the country-level analysis, we focus on 29 countries where IRCA admissions represented at least a third of total admissions also including refugees and the diversity visa, over 1983 to 2019. Section II.A of this Appendix describes these other admissions data in more detail.

Note that in forming these samples, we also restrict attention to metro areas with at least 20 registered Mexican LPRs in 1980 or countries with at least 20 registered LPRs (see Section I.C) and a legalization ratio of at least 0.1.

B. Immigrants Admitted to the United States

For the two much smaller legalization programs authorized by IRCA – the Cuban-Haitian Adjustment and Pre-1972 Arrivals programs – we obtain total admissions by country (29 sample countries) and metro area (for Mexicans only) across the 1987 to 2004 fiscal years from several sources: (1) for 1987 to 1997 from *Immigrants Admitted to the United States* microdata, available on ICPSR (all United States Department of Justice, Immigration and Naturalization Service, various years); and for 1998 and 2001-04 from the *Lawful Immigrants Files* version provided by the National Archives (Department of Homeland Security. Management Directorate.

---

[1] Statistics on IRCA admissions through fiscal year 2001, reported in Rytina (2002), show that nearly all IRCA admissions had occurred by the end of the 1992 fiscal year.

[2] For New England, we use New England County Metropolitan Areas (NECMAs). See June 30, 1999 definition at https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/historical-delineation-files.html.

[3] Because these metro areas are relatively large, the estimates are unaffected by the fact that county information is suppressed in the LAPS for applicants in counties with under 100,000 population (as of the 1990 census) or with fewer than 25 applications.

Office of Immigration Statistics, various years).[4] In table source notes, we refer to these files collectively as *Immigrants Admitted to the United States*. Like the LAPS, these data provide selected fields from anonymized records for Green Card admissions under all programs except the GLP and the SAW program. Because these data include detailed class of admission (identifying the relevant program), country of birth, and location within the U.S. at the time of admission, we are able to adjust the numerator of the legalization ratio for these two smaller legalization programs. We describe these data further in Section II of this Appendix.

C. Alien Address Reports

We obtain part of the denominator of the legalization ratio from *Alien Address Reports, [United States], 1980 Public Use File*, available at ICPSR. These public-use microdata consist of selected fields from anonymized records of registered aliens in the U.S. in 1980. LPRs are separately identified. These data were collected as part of the INS's alien address reporting program for 1980 and were used at the time to estimate unauthorized immigration in conjunction with the 1980 Census. The fields include country of birth and state and zip code of residence within the U.S., which we use to map to counties, and then to metro areas (see Section I.A of this Appendix).

D. Citizen Count

The denominator of the legalization ratio is the sum of the LPR count from I.C plus a count of citizens estimated from the 1980 Census PUMS (Ruggles et al., 2020). County groups in these data were matched to metro areas according to their 1999 definitions.

E. Descriptive Statistics

Table A2 shows how we arrived at the legalization ratio for each sampled metro area, by state. We show both the numerator (from sources I.A and I.B; column 2) and the denominator (from sources I.C, I.D; column 3) in addition to the ratio itself (column 1). We also show the share of IRCA admissions accounted for by that area (column 4). Table A7 shows how we arrived at the legalization ratio for each sampled country, by world region. We show both the numerator (from sources I.A and I.B; column 2) and the denominator (from sources I.C and I.D; column 3) in addition to the ratio itself (column 1). We also show the share of IRCA admissions accounted for by each country (column 4).

II. Outcomes Data: Immigrant Admissions

A. Immigrants Admitted to the United States

We calculate the first half of our country and metro-area panel on admissions by sponsor, relative type, and age from two sources: (1) *Immigrants Admitted to the United States* microdata, available on ICPSR, for fiscal years 1983-1997 and 1999-2000 (United States Department of Justice, Immigration and Naturalization Service, various years); (2) the National Archives version of this file for fiscal years 1998 and 2001-2004, the *Lawful Immigrant Files* (Department

[4] These visa categories are not separately identified in the 1999 and 2000 files, but their numbers are very small in 1998 and 2001.

of Homeland Security. Management Directorate. Office of Immigration Statistics, various years). In table source notes, we refer to these files collectively as *Immigrants Admitted to the United States* (1983-2004). These data provide selected fields from anonymized records for Green Card admissions under all programs except the GLP and the SAW program. These fields include detailed class of admission (identifying the relevant program), country of birth, and age and location within the U.S. at the time of admission.[5] In addition to identifying admissions under the Cuban-Haitian Adjustment and Pre-1972 Arrivals programs (see Section I.B of this Appendix), these data identify a variety of family-sponsorship visas, employer visas, diversity visas, and refugee visas.

We are constrained in what we can do with these data by the published tables that provide our main data source for fiscal years 2007 to 2019 (see section II.B). We categorize the family-sponsorship visas into two broad groups that align with what is available in later published data – e.g., a Green Card- sponsored category and a citizen-sponsored category. Likewise, among family-sponsored admissions overall, we are able to separate relatives into three categories – spouses and unmarried children of the sponsor, parents of the sponsor, and other relatives of the sponsor.

B. Office of Immigration Statistics Tables

Unfortunately for our study, publication of anonymized admissions microdata ceased after 2004. For the country-level analysis, we have collected tables for 2005 to 2019 from an online DHS database.[6] For the Mexican metro analysis, we relied on another online DHS database which is tabulated at the county level (for the largest immigrant destinations) from 2007 to 2019, which we further aggregate to the metropolitan area level.[7] So in addition to the constraints on these data noted in Section II.A, we lack data on Mexican admissions by metropolitan area for 2005-06, so we interpolate those years.

III. Outcomes Data: Total Arrivals

A. Data Sources and Construction

We estimated counts of recent immigrant arrivals ($total\ arrivals$ in equation (2)) by country, and for Mexico by U.S. metro area, from the 5% public-use microdata samples of the 1990 and 2000 Decennial Censuses (Ruggles, et al., 2020) and the public-use microdata samples of the 2006-2019 American Community Surveys. We focus on persons born in one of the 29 sample countries.[8] In calculating both the counts and the characteristics, we used survey-provided sampling weights.

Because the Census is not annual, we do not observe the size of all arrival cohorts at the time of arrival. We instead approximate it through extrapolation, taking advantage of the fact

[5] Location is recorded in different ways over time, e.g., initially and in 2001-04 as zip code and state and in 1999 and 2000 as metropolitan area. We convert all location information to metro areas (see Section I.A of this Appendix).
[6] https://www.dhs.gov/immigration-statistics/readingroom/LPR/LPR-by-major-class-and-country.
[7] https://www.dhs.gov/immigration-statistics/readingroom/LPR/LPRcounty.
[8] We exclude a small number of individuals born to U.S. citizens abroad.

that we observe each cohort at multiple points in time. Specifically, to create the data for the cross-metro Mexican analysis, we begin by estimating U.S. resident population counts of immigrant arrivals by survey year y, arrival year (or cohort) t, and metro area c, N_{cyt}. We normalize these counts by $legal_{c,1980}$ – the same denominator as is used for the legalization ratio. We then regress these normalized counts on a vector of area-by-arrival cohort fixed effects and a survey-specific effect of years in the U.S., $y - t$:

$$\frac{N_{cyt}}{legal_{c,1980}} = \eta_{ct} + \beta_1(y-t) + \beta_2(y-t) \times D(CENS) + v_{cyt}.$$

$D(CENS)$ is a dummy which indicates data are from 1990 or 2000 Census (rather than the American Community surveys). β_1 (or $\beta_1 + \beta_2$) is anticipated to be less than 1 to the extent that return migration or other forms of attrition shrink cohort sizes over time. To predict (normalized) cohort size at entry, we then fit of this model at three years in the U.S., i.e., $\frac{\widehat{N_{cyt}}}{legal_{c,1980}} = \hat{\eta}_{ct}$ when $y - t = 3$, roughly the midpoint of a five-year bin.

Indeed, arrival cohorts are not identified in single years in the 1990 Census: the available groupings are 1982-84, 1985-86, and 1987-90. We therefore also group 1980s arrivals in the 2000 Census and ACS 2006-2019 (where cohorts are reported in single arrival years) similarly: 1982-84, 1985-86, and 1987-89.[9] For these categories, we define "t" at the midpoint (that is, 1983, 1985.5, and 1988, respectively).

For the purposes of estimating the adjustment regression above, we drop those who arrived during the survey year (since full coverage of the year's arrival cohort will not be possible in a survey that takes place partway through the year) and only include cohorts within 18 years of the survey (so $1 \leq (y-t) \leq 18$). The latter restriction, for example, means only the 1990 and 2000 Censuses and the 2006 ACS give us observations on cohorts that arrived in the 1980s. We also can consider only cohorts up to t=2018 out of concern about how that the information in pandemic-era surveys may potentially differ.

Self-reported arrival cohorts are measured with a lot of error (e.g., Lubotsky 2007). To reduce noise, after the adjustment we further aggregated post-1990 arrivals into five-year intervals (1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2014, 2015-2018) by summing up the relevant $\hat{\eta}_{ct}$'s. We further inflated the counts to "five-year equivalent" intervals by scaling up each $\hat{\eta}_{ct}$ by 5/number of years in the interval (for example, 5/3 for 1982-4, 5/2 for 1985-86 and 5/4 for 2015-2018).

We also aggregate admissions A_{ct} – Mexican LPRs in area c in arrival cohort t – in the same way ($total\ admissions$ in equation (2)). That is, we aggregate A_{ct} into the same year intervals as the Census arrivals and adjust those to five-year equivalents as well. (Because of the missing 2005 and 2006 data, in particular, we adjust the 2005-2009 interval upwards by a factor of 5/3; we also have only 1983-1984 for the 1982-4 interval, so we adjust that upward by 5/2.)

[9] To be clear, 1990 Census defines the cohort as 1987-90, while later years we define the bin as 1987-89. We do this because the 1990 Census is taken in April, so most 1990 arrivals would not actually have been covered by the 1990 Census.

Finally, the difference $\hat{\eta}_{ct} - \frac{A_{ct}}{legal_{c,1980}}$ captures arrivals in all other immigrant categories ($other\ arrivals$ in equation (2)). We also follow the same procedure outlined to estimate adjusted data from the cross-country analysis substituting country for metro area for the "c" index.

B. Likely Unauthorized Immigrants

Figures A4 and A5 display the time-varying response of "likely unauthorized" immigrant arrivals. These data were constructed from Census and ACS data following a procedure like the one outlined in Borjas and Cassidy (2019). This procedure for identifying likely unauthorized immigrants is essentially a two-step "residual" approach. First, you identify non-citizens who are likely to be authorized based on working in a licensed occupation, being a veteran, receiving government benefits that are only available to authorized immigrants, or having a spouse with one of these attributes. Then, treat the remaining noncitizens as "likely unauthorized."[10] Our procedure differs in a few ways from Borjas and Cassidy (2019). As the Census has no information on Medicare and Medicaid use (used in the ACS), so we drop that part of their definition of "authorized" immigrants. SSI income is only available starting in 2000, so we do not use the 1990 Census for this. Finally, as we want separate estimates of counts by arrival cohort bins (described above) around the time of arrival, we follow the same linear adjustment procedure, described above, that we used for total arrivals.

IV. Other Data: Other Characteristics

A. Metro Area-Level Characteristics

We use tabulations of the 1980 Census 20% sample (Manson et al., 2020) to calculate the 1980 percent of a metro area's population who were Mexican. To calculate Mexicans admitted between 1983 and 1987 per legal Mexican in 1980, we use sources already described in I.B, I.C, and I.D above. Employment between 1980 and 1987 is calculated using *County Business Patterns* data (United States Bureau of the Census). We calculate the "Bartik" instrument for Mexican employment growth between 1980 and 2019 as follows:

$$\frac{\sum_o \frac{\Delta E_{o,-c}}{E_{o,-c,1980}} \widehat{Mex}_{oc,1980}}{Mex_{c,1980}},$$

where $\frac{\Delta E_{o,-c}}{E_{o,-c,1980}}$ is employment growth in occupation o in areas besides area c between 1980 and 2019 and $Mex_{c,1980}$ is the number of Mexicans in area c in 1980, and $\widehat{Mex}_{oc,1980} \equiv E_{oc,1980} \frac{Mex_{o,-c,1980}}{E_{o,-c,1980}}$ is the predicted number of Mexicans working in occupation o in area c in 1980 based on the Mexican share of that occupation outside the area, $\frac{Mex_{o,-c,1980}}{E_{o,-c,1980}}$, and the 1980 size of the occupation in that area, $E_{oc,1980}$. The idea of this measure is to leverage a combination of the local occupation mix and which occupations are growing fastest to predict which areas will

[10] This sort of approach was originally developed by the Pew Hispanic Center (see, for example, methodology section in Passel and Cohn (2018)), and modified by Borjas and Cassidy (2019).

become most attractive to Mexicans over the period of our study. All figures were computed using 1980 Census and 2019 ACS data from Ruggles et al. (2020).

B. Country-Level Characteristics

We used the 2000 Decennial Census (Ruggles et al., 2020) to approximate the naturalization rates of the IRCA cohort (entering 1971 to 1986) by country. For Mexicans in this cohort, we arrive at a naturalization rate of 35.6% – similar to Green Card holders entering the U.S. between 1979 and 1982 (35%), based on internal INS data through 2001. Like Rytina (2002), we also find a considerably higher naturalization rate for non-Mexicans – 55% in the Census versus 52% in the administrative data.

To calculate admissions between 1983 and 1987 per legal immigrant in 1980, we use sources already described in I.B, I.C, and I.D above. Upper income countries were identified using the United Nations World Development Indicators. Real exchange rates and population were computed using the Penn World Tables, version 10.0 (Feenstra et al., 2018). The population figures were normalized by the number of legal immigrants in 1980, previously described.

V. Tables

The data sources used in this project and their role in and use in this project are also summarized in Table A1.[11] Tables A2 and A7 report the raw data for the main cross-sectional variables used in the analysis (including the treatment) for the cross-metro area and -country analysis, respectively. Table A3 reports summary statistics for all of the variables used in the analysis. The remaining appendix tables are robustness checks for main analysis tables.

VI. References

A. Legalization Application Processing System Data.

Department of Justice. Immigration and Naturalization Service. Office of Strategic Planning. Statistics Division. 1991-1995. *1992 Legalization Summary Public Use Tape: Application Form I-687 Data File for California*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1315257.

Department of Justice. Immigration and Naturalization Service. Office of Strategic Planning. Statistics Division. 1991-1995. *1992 Legalization Summary Public Use Tape: Application Form I-687 Data File for States Other Than California*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1315263.

Department of Justice. Immigration and Naturalization Service. Office of Strategic Planning. Statistics Division. 1991-1995. *1992 Legalization Summary Public Use Tape: Application Form I-700 Data File*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1315265.

[11] We thank Cynthia Bansak for drafting this table for her discussion of our paper.

B. Immigrants Admitted to the United States.

----. *Immigrants Admitted to the United States, 1983*. Inter-university Consortium for Political and Social Research [distributor], 2010-04-02. https://doi.org/10.3886/ICPSR08963.v2

----. *Immigrants Admitted to the United States, 1984*. Inter-university Consortium for Political and Social Research [distributor], 2010-04-02. https://doi.org/10.3886/ICPSR08964.v2

----. *Immigrants Admitted to the United States, 1985*. Inter-university Consortium for Political and Social Research [distributor], 2010-04-07. https://doi.org/10.3886/ICPSR08965.v2

----. *Immigrants Admitted to the United States, 1986*. Inter-university Consortium for Political and Social Research [distributor], 2010-04-07. https://doi.org/10.3886/ICPSR08966.v2

----. *Immigrants Admitted to the United States, 1987*. Inter-university Consortium for Political and Social Research [distributor], 2010-04-19. https://doi.org/10.3886/ICPSR09268.v2

----. *Immigrants Admitted to the United States, 1988*. Inter-university Consortium for Political and Social Research [distributor], 2010-05-12. https://doi.org/10.3886/ICPSR09269.v2

----. *Immigrants Admitted to the United States, 1989*. Inter-university Consortium for Political and Social Research [distributor], 2010-05-13. https://doi.org/10.3886/ICPSR06161.v2

----. *Immigrants Admitted to the United States, 1990*. Inter-university Consortium for Political and Social Research [distributor], 2010-05-14. https://doi.org/10.3886/ICPSR06164.v2

----. *Immigrants Admitted to the United States, 1991*. Inter-university Consortium for Political and Social Research [distributor], 2010-08-24. https://doi.org/10.3886/ICPSR06165.v2

----. *Immigrants Admitted to the United States, 1992*. Inter-university Consortium for Political and Social Research [distributor], 2010-08-26. https://doi.org/10.3886/ICPSR06449.v2

----. *Immigrants Admitted to the United States, 1993-1995*. [distributor], 2007-01-17. https://doi.org/10.3886/ICPSR02267.v1

----. *Immigrants Admitted to the United States, 1996*. Inter-university Consortium for Political and Social Research [distributor], 2010-09-17. https://doi.org/10.3886/ICPSR02534.v2

----. *Immigrants Admitted to the United States, 1997*. Inter-university Consortium for Political and Social Research [distributor], 2007-11-05. https://doi.org/10.3886/ICPSR02955.v1

Department of Homeland Security. Management Directorate. Office of Immigration Statistics. *Lawful Immigrant Files, Fiscal Year 1998*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1302576.

United States Department of Justice. Immigration and Naturalization Service. *Immigrants Admitted to the United States, 1999*. [distributor], 2008-08-08. https://doi.org/10.3886/ICPSR03485.v2

----. *Immigrants Admitted to the United States, 2000*. [distributor], 2007-11-27. https://doi.org/10.3886/ICPSR03486.v2

Department of Homeland Security. Management Directorate. Office of Immigration Statistics. *Lawful Immigrant Files*, Fiscal Year 2001. Distributed by the United States National Archives, https://catalog.archives.gov/id/1302579

----. *Lawful Immigrant Files, Fiscal Year 2002*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1302580

----. *Lawful Immigrant Files, Fiscal Year 2003*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1302581

----. *Lawful Immigrant Files, Fiscal Year 2004*. Distributed by the United States National Archives, https://catalog.archives.gov/id/1302582

C. Alien Address Reports in 1980.

United States Department of Justice. Immigration and Naturalization Service. *Alien Address Reports, [United States]: 1980 Public Use File*. [distributor], 1992-02-16. https://doi.org/10.3886/ICPSR07998.v1

D. Census and American Community Survey

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles (2020). IPUMS National Historical Geographic Information System: Version 15.0 [dataset]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V15.0

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek (2020). IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V10.0

E. County Business Patterns

United States. Bureau of the Census. *County Business Patterns, 1980 [United States]: U.S. Summary, State, and County Data*. [distributor], 2006-01-12. https://doi.org/10.3886/ICPSR08142.v1

----. *County Business Patterns, 1987 [United States]: U.S. Summary, State, and County Data*. [distributor], 2006-01-12. https://doi.org/10.3886/ICPSR09381.v1

F. Penn World Tables

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182, available for download at www.ggdc.net/pwt

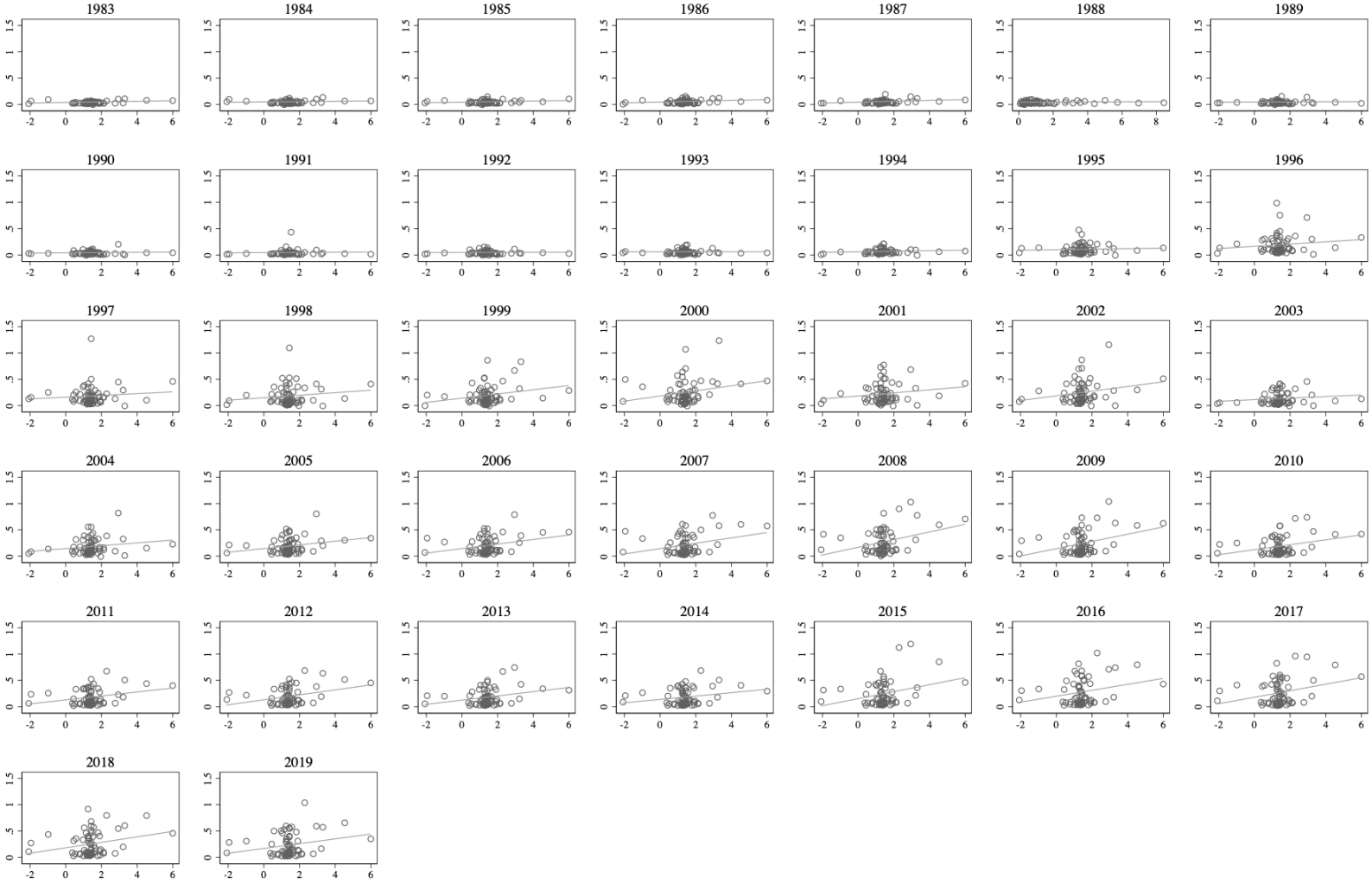
G. Other

Borjas, George J. and Hugh Cassidy. 2019. The wage penalty to undocumented immigration. *Labour Economics* 61.

Lubotsky, Darren. 2007. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy* 115(5): 820-867.

Passel, Jeffrey S. and D'Vera Cohn. 2018. "U.S. Unauthorized Immigrant Total Dips to Lowest Level in a Decade" Pew Research Center post dated November 27, 2018. https://www.pewresearch.org/hispanic/2018/11/27/u-s-unauthorized-immigrant-total-dips-to-lowest-level-in-a-decade/, accessed 20 December 2019.

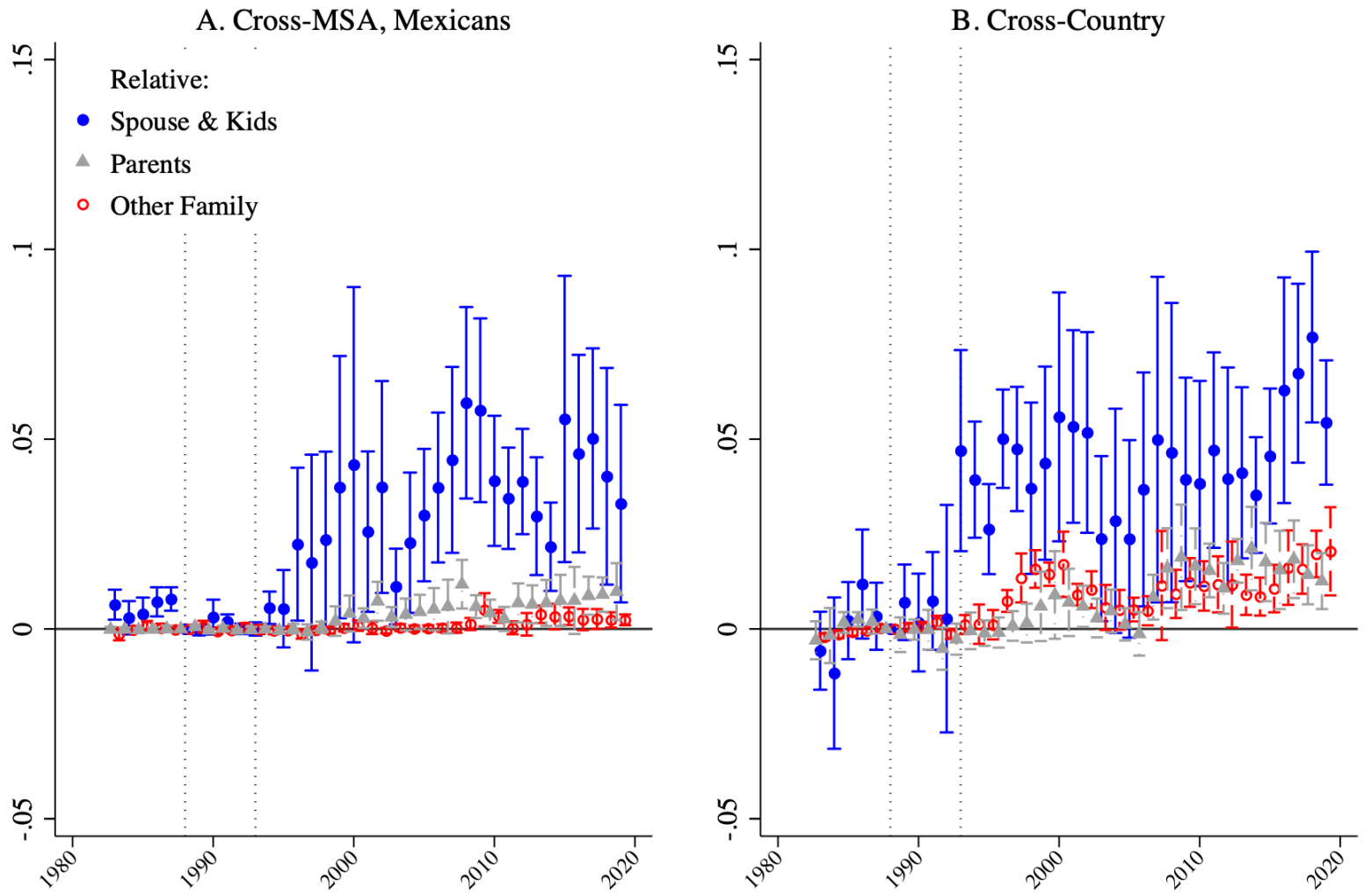
Figure A1. Cross-MSA Relationship between Mexican Family Admissions and Mexican Legalization Ratio, by Year



Sources: See Table 1 Panel A source notes for legalization ratio. Data for overall family admissions from *Immigrants Admitted to the United States* (FY 1983-2004) and https://www.dhs.gov/immigration-statistics/readingroom/LPR/LPRcounty (FY 2007-2019). Admissions for FY 2005 and 2006 are linearly interpolated. See Online Appendix A.

Notes: Thumbnail graphs are scatterplots between overall family admissions (y-axis) and the residual legalization ratio (x-axis). Residuals are from a regression of the legalization ratio on state fixed effects, to match our baseline specification. The slopes of the lines plotted thus match the points plotted for all family admissions in Figure 4 Panel A.

Figure A2. Response to IRCA Legalizations, by Family Relationship



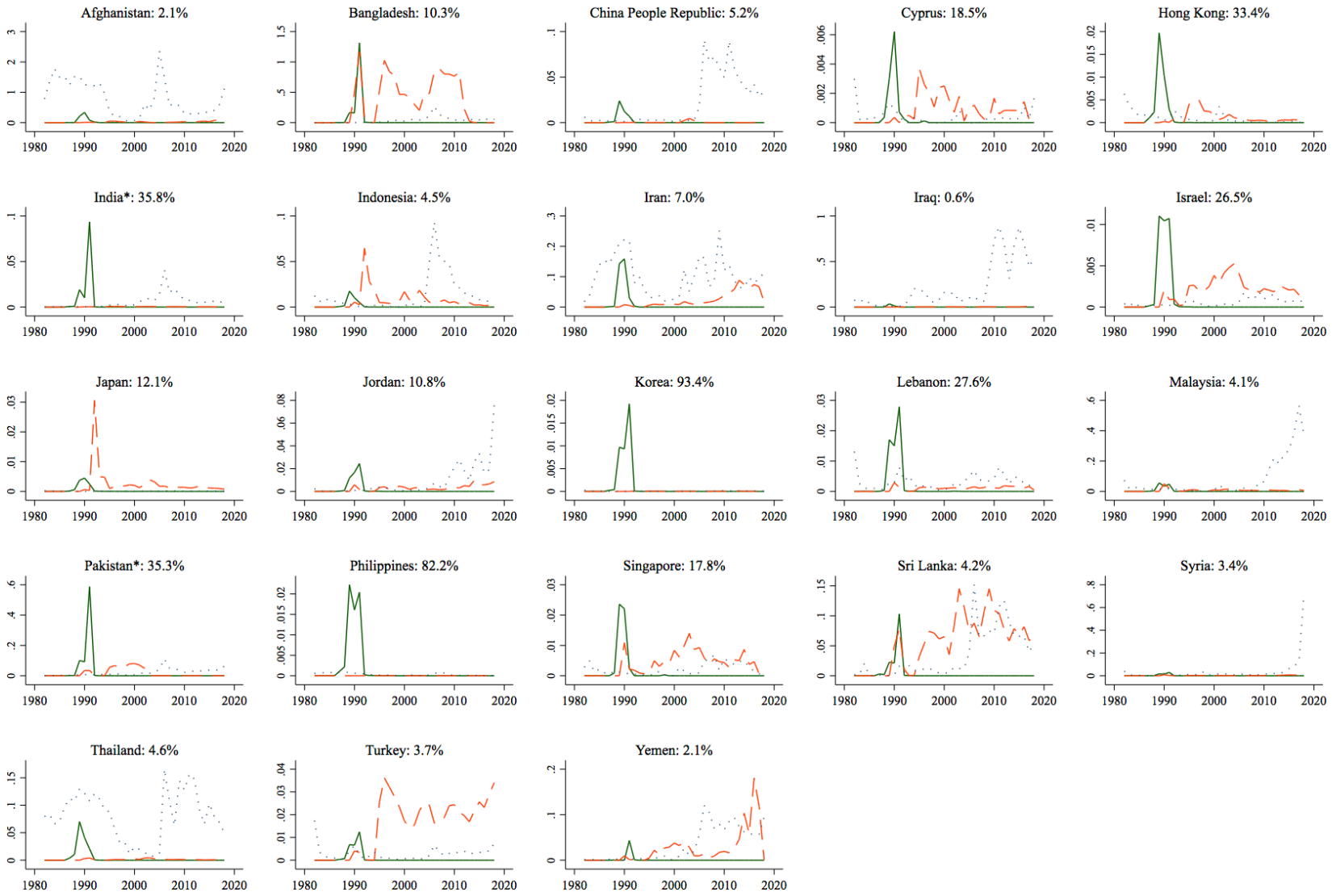
Sources: See Table 1 Panel A source notes for legalization ratio. Data on admissions by relative type are from *Immigrants Admitted to the United States* (FY 1983-2004) and https://www.dhs.gov/immigration-statistics/readingroom/LPR/LPRcounty for metro areas (FY 2007-2019) and https://www.dhs.gov/immigration-statistics/readingroom/LPR/LPR-by-major-class-and-country for countries (for FY 2005-2019). Admissions for FY 2005 and 2006 are linearly interpolated in the metro area analysis.

Notes: Panel A plots coefficients (with 90% confidence intervals) on the Mexican legalization ratio interacted with year dummies from a regression that also includes metro area and year-by-state fixed effects; the interaction between the legalization ratio and the dummy for 1988 is omitted. Estimation sample includes the 66 metro areas listed in Table A2. Regressions give each metro area equal weight, and standard errors are clustered on metro area. Panel B plots coefficients (with 90% confidence intervals) on the legalization ratio interacted with year dummies from a regression that also includes country and year-by-world region fixed effects; the interaction between the legalization ratio and the dummy for 1988 is omitted. Estimation sample includes the 29 countries listed in Table A7. Regressions give each country equal weight, and standard errors are clustered on country.

Figure A3a. Europe. IRCA Legalizations, Diversity, Refugees (w/IRCA's %)

Sources: Legalization Applications Processing System (LAPS) for IRCA legalizations and *Immigrants Admitted to the United States* (FY 1983-2004) and *Yearbook of Immigration Statistics* (FY 2005-2019) for remaining variables. See Online Appendix A.
Notes: Solid green = IRCA legalizations, dashed orange = diversity visas, dotted blue = refugees, per 1980 legal immigrant of that origin. *Country in sample.

Figure A3b. Asia. IRCA Legalizations, Diversity, Refugees (w/IRCA's %)

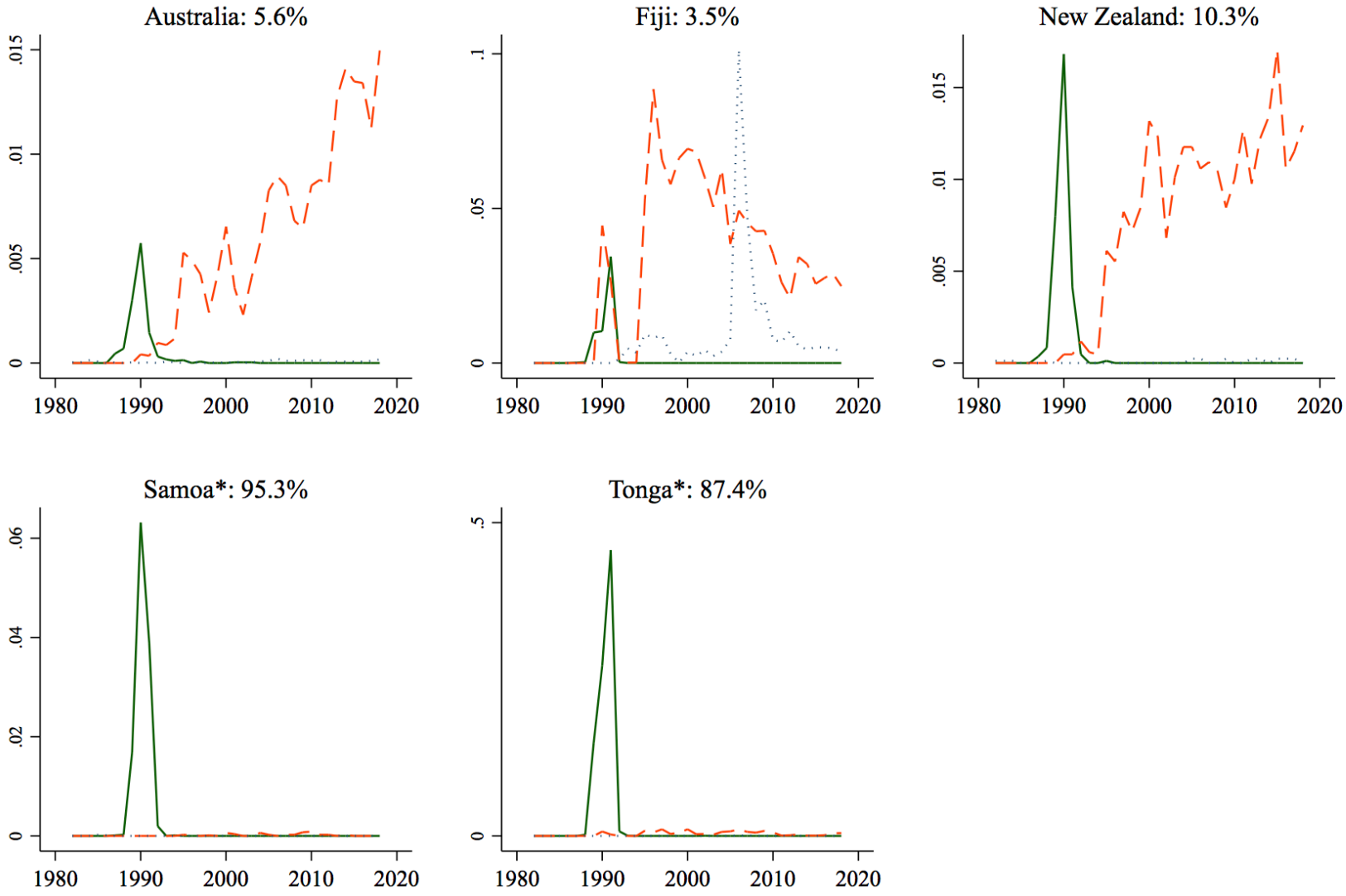


Sources: See Figure A3a.
Notes: Solid green = IRCA legalizations, dashed orange = diversity visas, dotted blue = refugees, per 1980 legal immigrant of that origin. *Country in sample.

Figure A3c. Africa. IRCA Legalizations, Diversity, Refugees (w/IRCA's %)

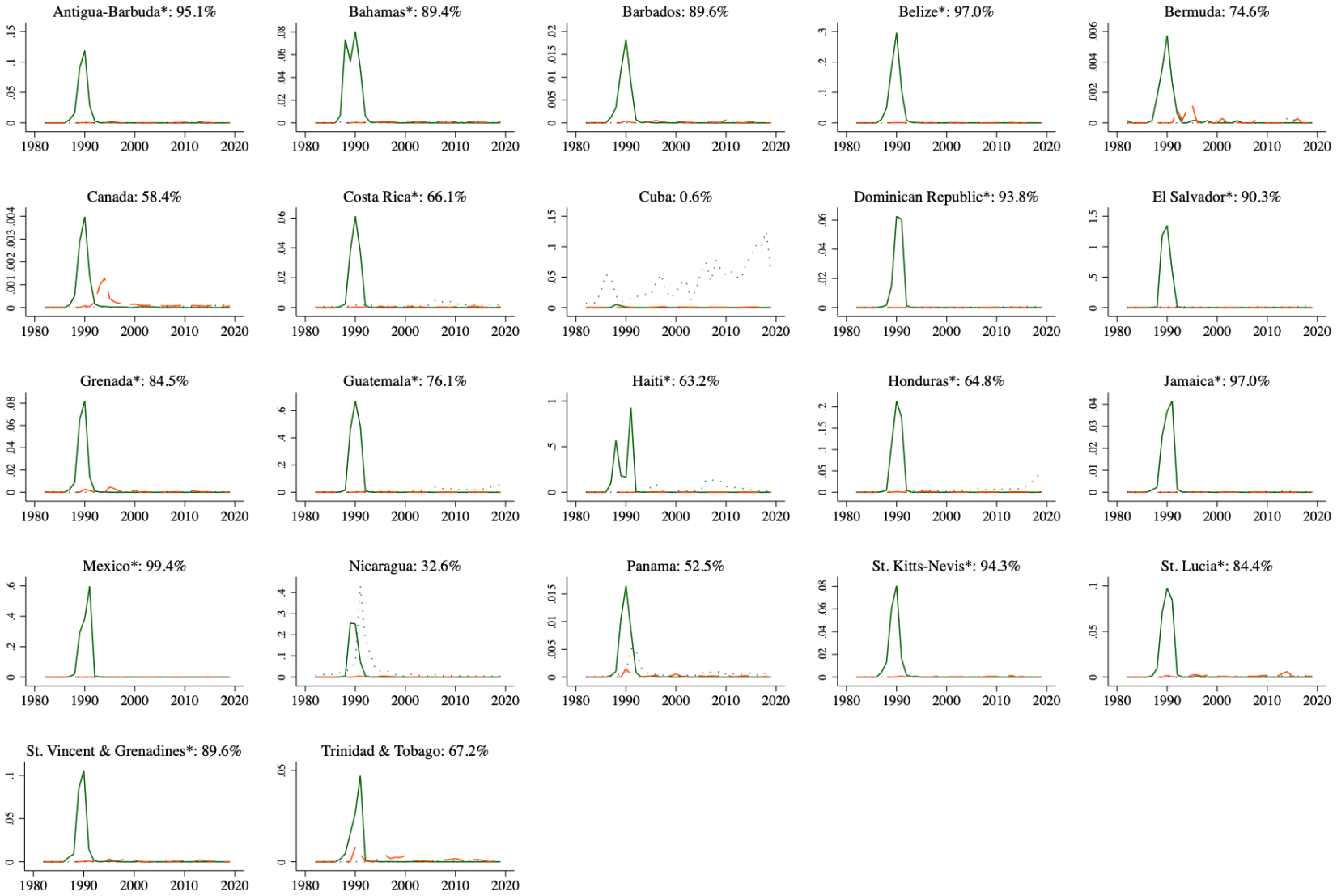
Sources: See Figure A3a.
Notes: Solid green = IRCA legalizations, dashed orange = diversity visas, dotted blue = refugees, per 1980 legal immigrant of that origin. *Country in sample.

Figure A3d. Oceania. IRCA Legalizations, Diversity, Refugees (w/IRCA's %)



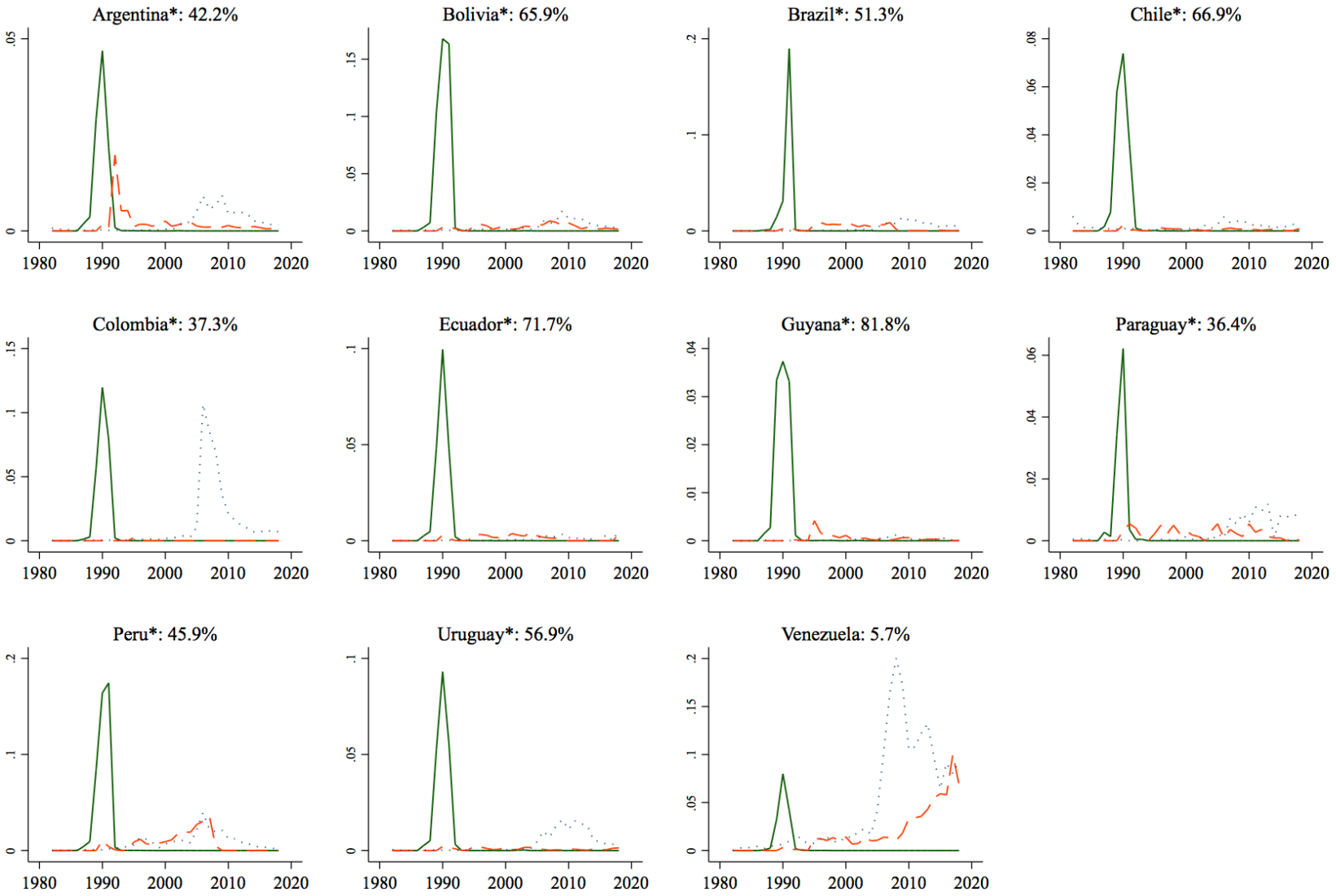
Sources: See Figure A3a.
Notes: Solid green = IRCA legalizations, dashed orange = diversity visas, dotted blue = refugees, per 1980 legal immigrant of that origin. *Country in sample.

Figure A3e. Central/North America. IRCA Legalizations, Diversity, Refugees (w/IRCA's %)



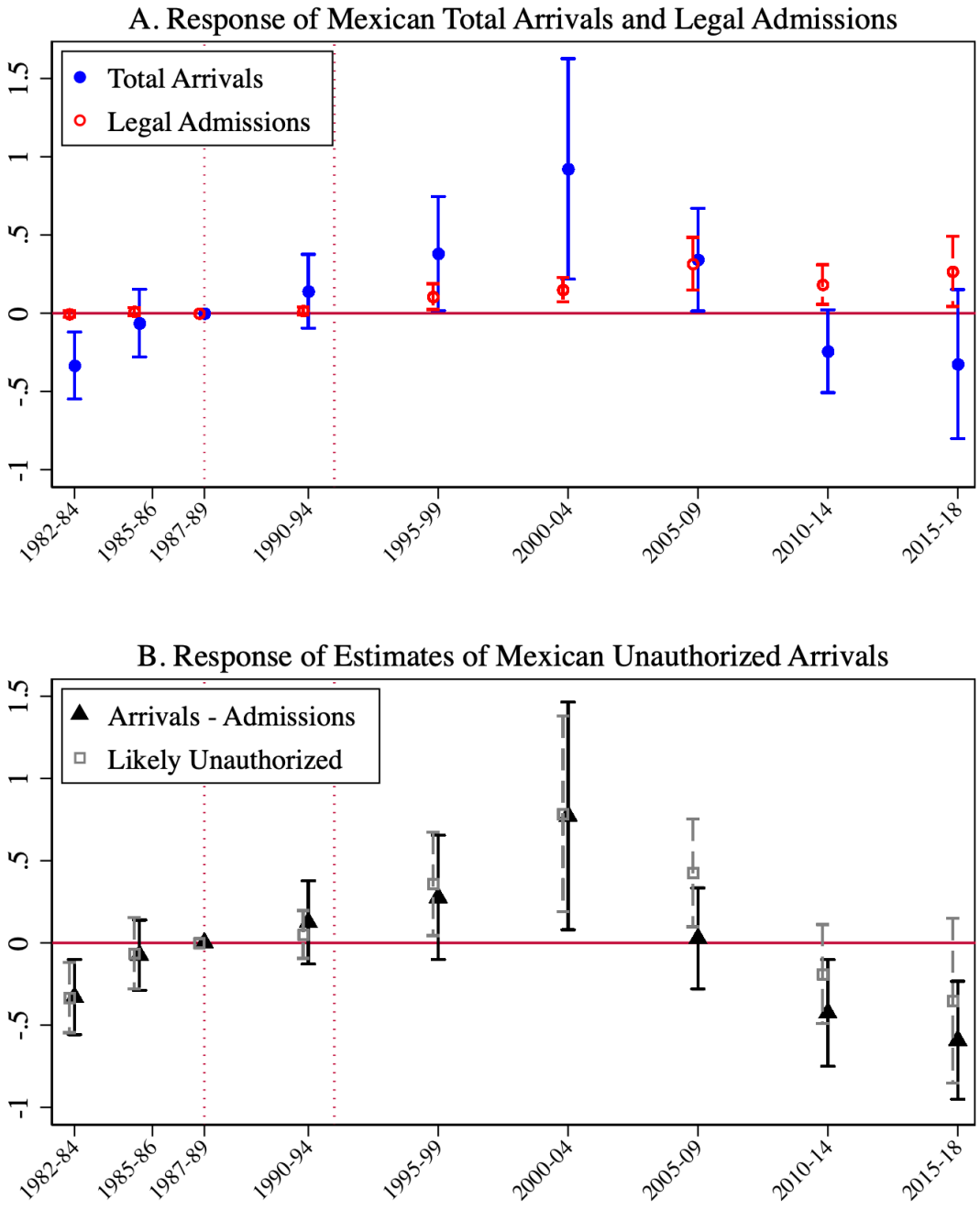
Sources: See Figure A3a.
Notes: Solid green = IRCA legalizations, dashed orange = diversity visas, dotted blue = refugees, per 1980 legal immigrant of that origin. *Country in sample.

Figure A3f. South America. IRCA Legalizations, Diversity, Refugees (w/IRCA's %)



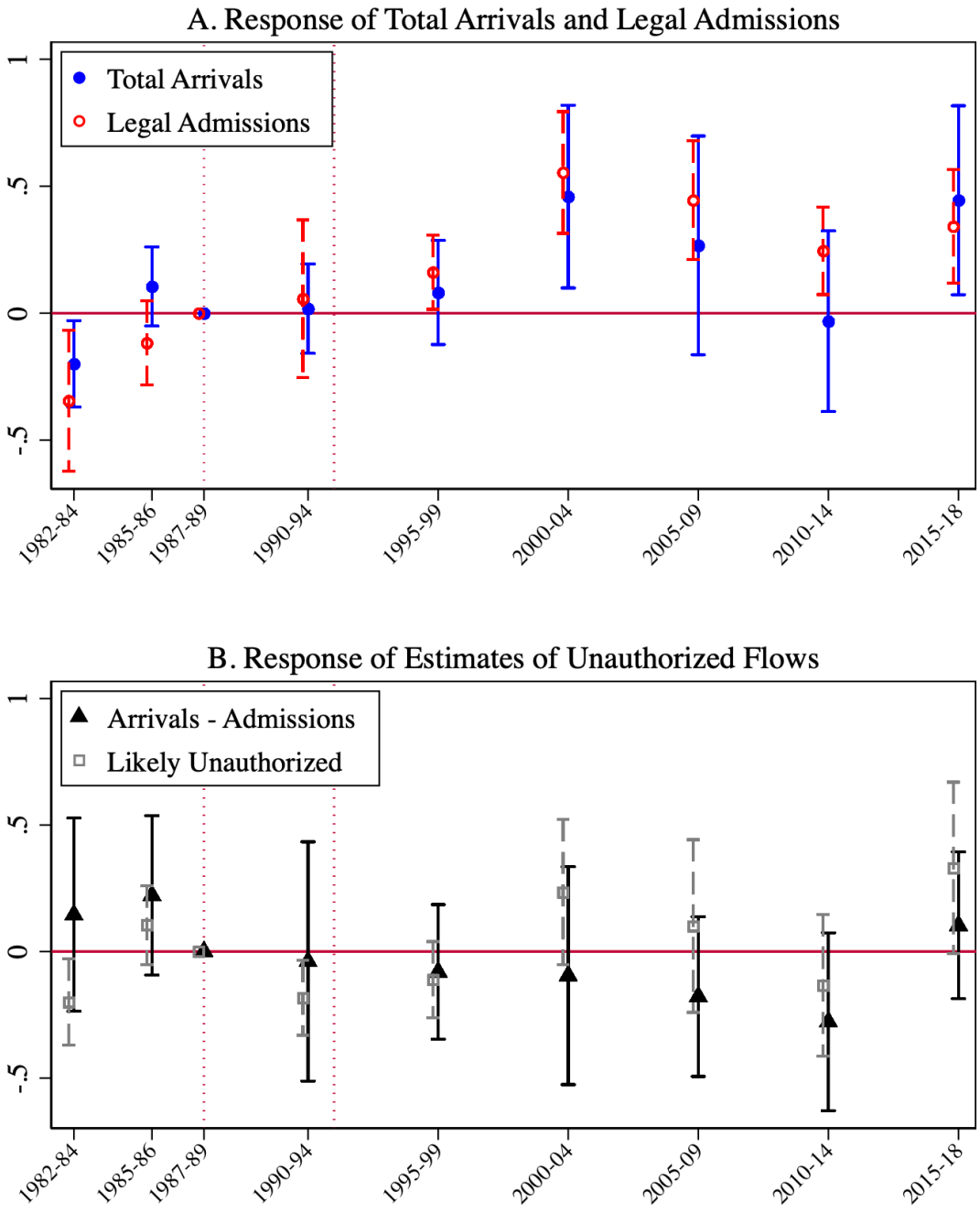
Sources: See Figure A3a.
Notes: Solid green = IRCA legalizations, dashed orange = diversity visas, dotted blue = refugees, per 1980 legal immigrant of that origin. *Country in sample.

Figure A4. Responses by Arrival Mode, Mexicans



Sources: American Community Survey and Census (Ruggles et al. 2020) for arrivals and administrative immigration sources for admissions and legalization ratio (see Figure A1). From a regression across 66 metropolitan areas, Panel A shows year-specific responses), relative to 1987-89, to the legalization ratio, of (1) total Mexican arrivals (in blue, measured in the Census/ACS) as well as (2) total legal Mexican Green Card admissions in the same bins (in red). Both have all been adjusted to five-year equivalent bins; the arrivals are also regression adjusted for return migration. Panel B shows the difference in these responses (sold lines) as well as the response of a measure of "likely unauthorized" workers defined similarly to Borjas and Cassidy (2019) using the Census/ACS (dashed lines, in grey).

Figure A5. Responses by Arrival Mode, Cross-Country Analysis



Sources: American Community Survey and Census (Ruggles et al. 2020) for arrivals and administrative immigration sources for admissions and legalization ratio (see Figure A1). From a regression across 29 countries, Panel A shows year-specific responses), relative to 1987-89, to the legalization ratio, of (1) total arrivals (in blue, measured in the Census/ACS) as well as (2) total legal Green Card admissions in the same bins (in red). Both have all been adjusted to five-year equivalent bins; the arrivals are also regression adjusted for return migration. Panel B shows the difference in these responses (sold lines) as well as the response of a measure of "likely unauthorized" workers defined similarly to Borjas and Cassidy (2019) using the Census/ACS (dashed lines, in grey).

Table A1. Summary of Data Sources

Type of Data and Source	Variable Description	Years covered	Analyses	Imputations
Outcomes: Legal Admissions to the U.S. or "LPR"s				
Immigrants Admitted to the United States	Immigrant admissions (LPRs)	FY1983 - 2004	Metro, Country	
DHS statistics tables	LPRs by State, County, Country of Birth, and Major Class of Admission (Top 200 Counties)	FY2007 - 2019	Metro	Interpolate 2005-6
DHS statistics tables	LPRs by Citizenship and Major Classes of Admission	FY2005 - 2019	Country	
Yearbook of Immigration Statistics	Immigrant admissions (Figure 1)	FY1940 - 2019	TS	
Legalized under IRCA				
Legalization Application Processing System data (LAPS)	IRCA applicant information	FY1988 - 1992	all	
Immigrants Admitted to the United States	Cuban-Haitian programs and pre-1972 arrivals	FY1988 - 1992	all	
Rytina (2002)	Legal status of IRCA applicants	FY1989 - 2002	TS	
Stock of all Legal U.S. Residents in 1980				
Alien Address Reports (INS)	Legalized immigrant population data (used in legalization ratio estimation)	1980	Metro, Country	
5% Public Use 1980 Decennial Census	Naturalized immigrant population	1980	Metro, Country	
Total (legal and unauthorized) arrivals				
Decennial Census	Total Immigrant arrivals (authorized and not)	1990, 2000	Metro, Country	Extrapolation to recent arrivals bins by year
American Community Survey	Total immigrant arrivals (authorized and not)	2006 - 2019	Metro, Country	"
Controls				
Public Use Decennial Census / American Community Survey	Bartik-style predicted Mexican employment growth	1980 - 2019 (Based on 1980)	Metro	
Tabulations of 20% count 1980 Decennial Census	Mexicans/Population	1980	Metro	
County Business Patterns	Employment growth 1983-1987	1983 - 1987	Metro	
Public Use 2000 Decennial Census	Share Naturalized among 1971-1986 arrivcals	2000, for 1971-1986 arrivals	Country	
Penn World Tables	Real exchange rate, growth in origin country population	1987 - 2018	Country	
UN World Development Indicators	Upper income country		Country	

Notes: TS = Used in time series (shown in some figures). See Online Appendix A text for further description of these sources.

Table A2. Treatment Variation and Characteristics of Mexicans: All Metro Areas, by State

State and Metro Area	**Treatment:** Legalization Ratio: (2)/(3)	Number Legalized by IRCA	Existing Legal Immi-grants, 1980	% of IRCA Legalizations	Characteristics Mexicans/Pop, %, 1980	Characteristics Emp Growth, %, 1980-87
	(1)	(2)	(3)	(4)	(5)	(6)
Arizona						
Phoenix-Mesa, AZ	0.78	18,248	23,519	0.90	1.81	44.5
Tucson, AZ	0.53	8,618	16,109	0.43	3.00	36.7
Yuma, AZ	0.98	9,737	9,929	0.48	12.11	36.7
California						
Bakersfield, CA	1.47	24,485	16,682	1.21	5.19	17.7
Los Angeles-Long Beach, CA	1.70	560,289	329,865	27.75	9.33	11.4
Merced, CA	2.02	12,593	6,228	0.62	7.83	20.4
Modesto, CA	1.35	12,423	9,183	0.62	4.35	27.1
Oakland, CA	0.44	10,142	23,232	0.50	1.69	28.4
Orange County, CA	2.68	108,593	40,546	5.38	4.50	35.2
Riverside-San Bernardino, CA	0.95	44,102	46,329	2.18	3.39	47.3
Sacramento, CA	0.28	2,725	9,862	0.13	1.10	40.0
Salinas, CA	1.35	21,841	16,171	1.08	8.87	19.9
San Diego, CA	1.22	83,744	68,912	4.15	4.67	37.9
San Francisco, CA	0.34	6,343	18,543	0.31	1.76	9.3
San Jose, CA	1.11	30,462	27,426	1.51	2.78	20.9
Santa Barbara-Santa Maria-Lompoc, CA	1.84	19,538	10,642	0.97	4.36	21.0
Santa Rosa, CA	3.13	8,362	2,675	0.41	1.43	46.1
Stockton-Lodi, CA	1.18	15,402	13,083	0.76	3.97	29.0
Vallejo-Fairfield-Napa, CA	0.71	3,130	4,426	0.16	1.68	38.8
Ventura, CA	0.91	25,347	27,948	1.26	6.55	50.8
Visalia-Tulare-Porterville, CA	2.04	25,424	12,467	1.26	7.60	14.4
Yolo, CA	0.69	3,148	4,558	0.16	4.39	52.2
Colorado						
Colorado Springs, CO	0.60	197	326	0.01	0.14	44.9
Denver, CO	0.34	2,105	6,215	0.10	0.67	11.7

Table A2. Treatment Variation and Characteristics of Mexicans: All Metro Areas, by State (continued)

Region and Country	**Treatment:** Legalization Ratio: (2)/(3)	Number Legalized by IRCA	Existing Legal Immi-grants, 1980	% of IRCA Legalizations	Characteristics Mexican Pop %, 1980	Characteristics Emp Growth, %, 1980-87
	(1)	(2)	(3)	(4)	(5)	(6)
Connecticut						
New Haven, CT	0.69	233	338	0.01	0.03	15.8
Florida						
Fort Lauderdale, FL	3.38	1,462	432	0.07	0.05	30.6
Fort Myers-Cape Coral, FL	6.94	1,958	282	0.10	0.15	56.5
Fort Pierce-Port St. Lucie, FL	5.72	898	157	0.04	0.51	65.0
Lakeland-Winter Haven, FL	8.41	4,162	495	0.21	0.19	16.4
Melbourne, FL	0.34	70	208	0.00	0.03	47.1
Naples, FL	4.37	5,428	1,241	0.27	1.32	66.8
Orlando, FL	0.47	423	909	0.02	0.19	63.5
Sarasota-Bradenton, FL	3.55	1,286	362	0.06	0.15	46.6
Tampa-St. Petersburg, FL	1.43	1,820	1,272	0.09	0.08	44.5
West Palm Beach, FL	3.81	4,103	1,077	0.20	0.12	58.3
Hawaii						
Honolulu, HI	0.27	126	472	0.01	0.08	12.7
Illinois						
Chicago, IL	0.20	20,695	101,396	1.03	2.23	2.8
Massachusetts						
Boston, MA	0.11	95	854	0.00	0.02	22.0
Springfield, MA	0.30	8	25	0.00	0.02	9.8
Nevada						
Reno, NV	5.00	3,377	676	0.17	0.71	20.3

Table A2. Treatment Variation and Characteristics of Mexicans: All Metro Areas, by State (continued)

Region and Country	**Treatment:** Legalization Ratio: (2)/(3)	Number Legalized by IRCA	Existing Legal Immi-grants, 1980	% of IRCA Legalizations	Characteristics: Mexican Pop %, 1980	Characteristics: Emp Growth, %, 1980-87
	(1)	(2)	(3)	(4)	(5)	(6)
New Jersey						
Bergen-Passaic, NJ	1.04	654	629	0.03	0.08	17.7
Jersey City, NJ	1.03	331	320	0.02	0.06	10.6
Middlesex-Somerset, NJ	1.03	192	187	0.01	0.04	35.8
Monmouth-Ocean, NJ	2.72	300	110	0.01	0.02	45.4
Newark, NJ	0.22	89	406	0.00	0.03	13.6
New York						
Buffalo-Niagara Falls, NY	0.15	31	207	0.00	0.02	-0.8
Nassau-Suffolk, NY	0.45	343	758	0.02	0.03	31.7
New York, NY	0.32	1,729	5,400	0.09	0.10	9.0
Oregon						
Portland, OR	1.22	1,911	1,572	0.09	0.20	6.1
Pennsylvania						
Allentown, PA	0.27	37	139	0.00	0.02	3.0
Lancaster, PA	1.03	65	63	0.00	0.03	17.7
Philadelphia, PA	0.39	380	969	0.02	0.03	15.0
Texas						
Brazoria, TX	0.91	2,315	2,555	0.11	1.87	-14.4
Brownsville, TX	0.34	12,909	37,900	0.64	16.70	9.5
El Paso, TX	0.33	27,884	84,284	1.38	17.31	11.5
Houston, TX	0.42	28,352	67,082	1.40	3.29	-0.4
Laredo, TX	0.18	4,569	25,867	0.23	18.83	4.2
McAllen, TX	0.43	24,858	57,874	1.23	18.01	27.7

Table A2. Treatment Variation and Characteristics of Mexicans: All Metro Areas, by State (continued)

Region and Country	**Treatment:** Legalization Ratio: (2)/(3)	Number Legalized by IRCA	Existing Legal Immi-grants, 1980	% of IRCA Legalizations	Characteristics Mexican Pop %, 1980	Characteristics Emp Growth, %, 1980-87
	(1)	(2)	(3)	(4)	(5)	(6)
Texas (continued)						
San Antonio, TX	0.35	16,835	48,547	0.83	4.62	27.4
Utah						
Provo-Orem, UT	2.17	721	332	0.04	0.32	3.9
Salt Lake City-Ogden, UT	0.46	782	1719	0.04	0.28	14.7
Washington						
Seattle-Bellevue-Everett, WA	0.86	1,169	1367	0.06	0.10	19.3
Tacoma, WA	1.64	498	304	0.02	0.10	17.5
Wisconsin						
Madison, WI	0.73	119	164	0.01	0.09	21.9

Sources: Columns 2 and 4: Legalization Applications Processing System (LAPS) (for the SAW program and the GLP) and *Immigrants Admitted to the United States* (FY 1987-2004) (for two smaller IRCA programs). Column 3: *Alien Address Reports, [United States], 1980 Public Use File* (for LPRs) and 1980 Census PUMS (Ruggles et al., 2020) (for citizens). Column 5: 1980 Census tabulations (Manson et al., 2020). Column 6: County Business Patterns. See text of Online Appendix A.

Table A3. Descriptive Statistics of Key Variables, 1983-2019

	Metro Area (1)	Country (2)		Metro Area (3)	Country (4)
1. Treatment			*Other Major Categories (Continued)*		
Legalization Ratio	1.406	0.525	Refugees	0.001	0.005
Ratio (treatment)	(1.643)	(0.681)		(0.004)	(0.013)
			Diversity	0.000	0.002
2. Mexican Legal Admissions (all per 1980 Mexican Citizens+Permanent Residents)				(0.000)	(0.007)
Overall Family Sponsored	0.156	0.148	3. Controls		
	(0.177)	(0.110)	Mexicans/Population, 1980	0.029	
By Family Sponsorship Type				(0.046)	
Green-Card Sponsored	0.039	0.031	Mexicans Admitted, 1983-87	3.845	
	(0.061)	(0.036)	/Legal Mexicans, 1980	(9.331)	
Citizen-Sponsored	0.117	0.118	Employment Growth, 1980-87	0.258	
	(0.139)	(0.094)		(0.179)	
By Relative Type			Mex Emp Growth, 1980-2019	5.452	
Spouses and Kids[a]	0.125	0.100	predicted from 1980 Occ Mix	(8.824)	
	(0.143)	(0.072)	Admissions 1983-87/1980		0.980
Parents	0.022	0.019	Legal Immigrants		(0.571)
	(0.035)	(0.019)	Upper Income Country		0.310
Other Relatives[b]	0.008	0.030			(0.463)
	(0.011)	(0.034)	Real Exchange Rate		2.462
Other Major Categories					(3.603)
Employer-Sponsored	0.018	0.022	Origin Country Population		0.970
	(0.051)	(0.036)	/1K Legal Imms, 1980		(1.990)
Observations (cells)	2,310	1,073		2,310	1,073
Countries[c]	Mexico Only	29		Mexico Only	29
Metro Areas	66	(national)		66	(national)
Years	37	37		37	37

Notes: Table shows mean of referenced variable, with standard deviation in parentheses underneath.
[a] Sum of citizen-sponsored spouses and minor children and Green Card-sponsored spouses, minor children, and unmarried children.
[b] Siblings, married children, and citizen-sponsored unmarried adult children.
[c] In Panel 3, only 27 of the 29 total countries have the controls available.

Table A4. Cross-Metro Area Treatment Variation, Additional Correlates

Panel A. Top MSAs on Mexican Legalization Ratio

	Metro Area	Legalization Ratio	% Legalized Under SAW	% of Apps Accepted
		(1)	(2)	(3)
1	Lakeland-Winter Haven, FL	8.4	86.8	93.0
2	Fort Myers-Cape Coral, FL	6.9	88.4	94.1
3	Fort Pierce-Port St. Lucie, FL	5.7	93.0	94.3
4	Reno, NV	5.0	45.5	84.3
5	Naples, FL	4.4	91.0	95.5
6	West Palm Beach-Boca Raton, FL	3.8	81.9	93.1
7	Sarasota-Bradenton, FL	3.6	87.1	95.4
8	Fort Lauderdale, FL	3.4	67.0	89.9
9	Santa Rosa, CA	3.1	73.8	91.4
10	Monmouth-Ocean, NJ	2.7	66.4	79.3
16	Los Angeles-Long Beach, CA	1.7	21.7	87.0

Panel B. Correlations with the Legalization Ratio

Characteristic	Mean	Regressions on Leg. Ratio	
	(1)	(2)	(3)
% Legalized under SAW	58.61	5.714	2.691
		(1.202)	(1.172)
% of Applications Accepted.	85.70	1.506	0.188
		(0.335)	(0.248)
State Effects?		No	Yes

Sources: Panel A column 1 numerator: Legalization Applications Processing System (LAPS) (for the SAW program and the GLP) and *Immigrants Admitted to the United States* (FY 1987-2004) (for two smaller IRCA programs). Panel A column 1 denominator: *Alien Address Reports, [United States], 1980 Public Use File* (for LPRs) and 1980 Census PUMS (Ruggles et al., 2020) (for citizens). Panel A columns 2 and 3 and Panel B: LAPS microdata.

Notes: Unit of observation is a metro area. The legalization ratio in Panel A gives the number of Mexican immigrants granted permanent residence by IRCA who listed that metro area as their intended residence, divided by the number of Mexican citizens and LPRs in that metro area in 1980. Columns 2 and 3 of Panel B show the coefficient from a regression of the variable listed on the legalization ratio; the regression in column 3 also includes dummies for the state in which the majority of the metro area's population resided in 1986.

Table A5. Impact of Mexican IRCA Legalizations by IRCA Program

Panel A: Balance Test

	SAW	GLP + Other	P-value on joint sig.
	(1)	(2)	(3)
Mexicans/Population, 1980	0.00	-0.01	0.579
	(0.00)	(0.02)	
Mexicans Admitted, 1983-87	0.57	1.07	0.384
/Legal Mexicans, 1980	(0.58)	(1.48)	
Employment Growth, 1980-87	0.00	-0.10	0.517
	(0.03)	(0.09)	
Mex Emp Growth, 1980-2019	0.06	2.61	0.649
predicted from 1980 Occ Mix	(0.75)	(2.81)	

Panel B: Long-Run Responses

	SAW	GL+Other	P-value on difference
	(1)	(2)	(3)
Overall Family Sponsored	1.06	0.85	0.856
	(0.30)	(1.03)	
By Family Sponsorship Type			
Green-Card Sponsored	0.57	0.01	0.046
	(0.11)	(0.25)	
Citizen-Sponsored	0.49	0.84	0.712
	(0.22)	(0.83)	
By Relative Type			
Spouses and Kids[a]	0.90	0.72	0.846
	(0.24)	(0.84)	
Parents	0.11	0.19	0.700
	(0.06)	(0.17)	
Other Relatives[b]	0.05	-0.06	0.138
	(0.02)	(0.06)	
Other Major Categories			
Employer-Sponsored	-0.09	0.80	0.258
	(0.14)	(0.67)	
Refugees	-0.01	0.05	0.381
	(0.02)	(0.06)	
Diversity	0.00	0.00	0.273
	(0.00)	(0.00)	

Sources: See notes to Table 1 for Panel A sources. Data on admissions by type (Panel B) from *Immigrants Admitted to the United States* (FY 1983-2004) and https://www.dhs.gov/immigration-statistics/readingroom/LPR/LPRcounty (FY 2007-2019). Admissions for FY 2005 and 2006 are linearly interpolated. See Online Appendix A.

Notes: Panel A shows coefficients and standard errors from a regression of the variable listed on the number of Mexican-born immigrants legalized under IRCA's SAW program (column 1) and under other IRCA legalization programs (column 2), each divided by $legal_{c,1980}$. Panel B gives the sum of post-1988 coefficients on the same two SAW and GLP variables interacted with dummies for year from a regression that also includes controls for metro area and state x year fixed effects. Standard errors in these regressions are clustered on metro area, and standard errors in parentheses are calculated using the delta method.

[a] Citizen-sponsored spouses and minor kids + Green Card-sponsored spouses, minor children and unmarried children.

[b] Siblings, married children, and citizen-sponsored unmarried adult children.

Table A6. Cross-Country Treatment Variation

Panel A. Top Countries on Legalization Ratio

	Country	Legalization Ratio: (2)/(3)	Legalized by IRCA	Legal Immigrants, 1980	% of all Legalizations
		(1)	(2)	(3)	(4)
4	Mexico	1.30	2,019,353	1,548,438	72.2
	Other 28 countries in sample		*512,056*	*1,156,230*	18.3
1	El Salvador	3.17	151,880	47,913	5.4
2	Haiti	1.95	88,284	45,209	3.2
3	Guatemala	1.64	63,663	38,742	2.3
5	Tonga	0.89	3,186	3,593	0.1
6	Pakistan	0.79	17,009	21,654	0.6
7	Belize	0.66	6,035	9,155	0.2
8	Honduras	0.51	16,055	31,422	0.6
9	Bolivia	0.45	4,337	9,666	0.2
10	Peru	0.44	18,264	41,522	0.7

Panel B. Balance Test: Correlates of the Legalization Ratio

	Characteristic	Mean	Regressed on Leg. Ratio		Reverse
		(1)	(2)	(3)	(4)
(a)	<60% of 1971-1986 Arrivals Naturalized by 2000	0.483	0.190	0.202	
			(0.108)	(0.120)	
(b)	Admissions 1983-87/ 1980 Legal Immigrants	0.980	0.126	0.0162	-0.0428
			(0.102)	(0.127)	(0.231)
(c)	Upper Income Country	0.310	-0.137	-0.143	-0.337
			(0.0783)	(0.0874)	(0.264)
(d)	Missing Country Controls (in (e) and (f))	0.0690	-0.00315	0.00412	
			(0.0428)	(0.0325)	
(e)	Δln(Real Exchange Rate), 1987-2018	2.930	-0.874	0.238	0.0085
			(0.554)	(0.357)	(0.0200)
(f)	Δ Country Pop, 1987-2018 /1K Legal Imms, 1980	0.537	-0.0828	0.0811	0.127
			(0.166)	(0.0689)	(0.117)
	Dummy Controls: North and South America		No	Yes	Yes
	F-stat				0.934

Sources: Panel A columns 2 and 4: Legalization Applications Processing System (LAPS) (for the SAW program and the GLP) and *Immigrants Admitted to the United States* (FY 1987-2004) (for two smaller IRVA programs). Panel A column 3: *Alien Address Reports, [United States], 1980 Public Use File* (for LPRs) and 1980 Census PUMS (Ruggles et al., 2020) (for citizens). Panel B row a: 2000 Census PUMS (Ruggles et al, 2020). Panel B row b: *Immigrants Admitted to the United States*, FY 1983-87 (numerator) and Panel A column 3 sources (denominator). Panel B row c: UN World Development Indicators. Panel B rows d, e, and f: Penn World Tables 10.0 (Feenstra et al., 2018). See Online Appendix A.

Notes: Unit of observation is an origin country. The legalization ratio in Panel A gives the number of immigrants from the country granted permanent residence by IRCA, divided by the number of citizens and LPRs from that country in 1980. Columns 2 and 3 of Panel B show the coefficient from a regression of the variable listed on the legalization ratio; the regression in column 3 also includes dummies for world region. Column 4 of Panel B shows the slope coefficients from a multivariate regression of the legalization ratio on the variables listed plus world region dummies; the F-stat is on the joint significance of the variables listed.

Table A7. Treatment Variation and Characteristics: All Countries, by Region

Region and Country	**Treatment:** Legalization Ratio: (2)/(3)	Number Legalized by IRCA	Existing Legal Immi-grants, 1980	% of IRCA Legalizations	Characteristics: Estimated % naturalized	Characteristics: Upper Income Country?
	(1)	(2)	(3)	(4)	(5)	(6)
1. Europe						
(none)						
2. Asia						
India	0.13	20,906	167,896	0.75	72.59	0
Pakistan	0.79	17,009	21,654	0.61	75.57	0
3. Africa						
(none)						
4. Pacific						
Samoa	0.12	994	8,186	0.04	65.46	0
Tonga	0.89	3,186	3,593	0.11	45.79	0
5. North America and Caribbean						
Antigua and Barbuda	0.26	1,268	4,808	0.05	68.34	1
The Bahamas	0.27	2,897	10,712	0.10	48.42	0
Belize	0.66	6,035	9,155	0.22	57.39	0
Costa Rica	0.14	3,363	23,882	0.12	59.08	1
Dominican Republic	0.14	23,982	169,257	0.86	50.33	0
El Salvador	3.17	151,880	47,913	5.43	41.14	0
Grenada	0.17	921	5,300	0.03	70.43	1
Guatemala	1.64	63,663	38,742	2.28	41.89	0
Haiti	1.95	88,284	45,209	3.16	59.83	0
Honduras	0.51	16,055	31,422	0.57	48.83	0
Jamaica	0.11	17,257	158,284	0.62	66.51	0
Mexico	1.30	2,019,353	1,548,438	72.20	35.61	1

Table A7. Treatment Variation and Characteristics: All Countries, by Region (continued)

Region and Country	**Treatment:** Legalization Ratio: (2)/(3)	Number Legalized by IRCA	Existing Legal Immi-grants, 1980	% of IRCA Legalizations	Characteristics Estimated % naturalized	Characteristics Upper Income Country?
	(1)	(2)	(3)	(4)	(5)	(6)
5. North America and Caribbean (cont'd)						
St. Kitts and Nevis	0.18	629	3,554	0.02	61.81	0
St. Lucia	0.27	619	2,309	0.02	67.65	1
St. Vincent & Grenadines	0.22	716	3,219	0.03	67.61	0
6. South America						
Argentina	0.10	5,619	53,804	0.20	60.27	1
Bolivia	0.45	4,337	9,666	0.16	57.83	0
Brazil	0.24	6,956	29,027	0.25	46.87	1
Chile	0.18	4,647	25,891	0.17	57.91	1
Colombia	0.26	30,941	118,215	1.11	61.44	0
Ecuador	0.21	15,274	74,392	0.55	52.26	0
Guyana	0.11	3,990	36,391	0.14	76.84	0
Paraguay	0.11	230	2,188	0.01	54.53	0
Peru	0.44	18,264	41,522	0.65	59.73	0
Uruguay	0.21	2,134	10,039	0.08	63.44	1

Sources: Columns 2 and 4: Legalization Applications Processing System (LAPS) (for the SAW program and the GLP) and *Immigrants Admitted to the United States* (FY 1987-2004) (for two smaller IRCA programs). Column 3: *Alien Address Reports, [United States], 1980 Public Use File* (for LPRs) and 1980 Census PUMS (Ruggles et al., 2020) (for citizens). Column 5: 2000 Census PUMS (Ruggles et al., 2020). Column 6: United Nations World Development Indicators. See text of Online Appendix A.
Notes: The naturalization rate is the percent of 1971-86 arrivals how were citizens as of the 2000 Census.

Appendix B: Derivation of the Main Estimation Equation

As discussed, our main empirical approach exploits variation in the timing of IRCA and variation across metropolitan areas in the intensity of IRCA as a legalization shock. To understand the specifications we ultimately estimate, it is helpful to outline a stylized model.

I. Baseline model

Let a_{cst} represent immigrants from Mexico admitted (with a Green Card) to metro area c in state s in year t.[1] We begin by modeling a_{cst} as a function of a metro area fixed effect, $\tilde{\delta}_c$, with deviations subsequent to IRCA (after 1988) proportional to the number of IRCA Green Cards issued to Mexicans in c, $lpr_{c,IRCA}$. That is,

$$a_{cst} = \tilde{\delta}_c + \sum_{\tau>1988} \theta_\tau D_t^\tau lpr_{c,IRCA} + u_{cst}, \tag{B.0}$$

where the D_t^τ represent an exhaustive set of indicator variables for all years after 1988 and u_{cst} is an error term, capturing other area-by-time varying determinants of Mexican admissions, including various "push" and "pull" factors. The coefficients of interest are the θ_τ's. With annual data, for example, θ_{2000} would be the difference in Mexican admissions to c between 1988 and 2000, on average, for every Mexican IRCA LPR in c.

II. Modifications

While this model is intuitively appealing, we think it necessary to modify in several ways to produce credible estimates of the θ_τ's.

A. Modification 1: Other Sources of Sponsorship

First, IRCA LPRs were not the only immigrants capable of sponsoring new LPRs through family linkages in the 1990s and beyond; pre-existing LPRs and citizens were capable of sponsoring family members as well. Though it may not be the case either that these other legal immigrants accelerated their sponsorship in the 1990s, or that these stocks of other legal immigrants are even all that correlated with $lpr_{c,IRCA}$, let's allow for this possibility:

$$a_{cst} = \widetilde{\delta_c} + \sum_{\tau>1988}\left(\theta_\tau D_t^\tau lpr_{c,IRCA} + \gamma_\tau D_t^\tau legal_{c,1988}\right) + \tilde{e}_{cst}, \tag{B.1}$$

where $legal_{c,1988}$ represents the stock of LPRs and citizens (combined) from country c in 1988 before IRCA. Model (B.1) thus adjusts $lpr_{c,IRCA}$ and a_{cst} for other ways in which a_{cst} may change over time.

B. Modification 2: Scaling

Even with modification 1, the model is susceptible to influence from outliers due to regional concentrations of immigrants (e.g., areas, like Los Angeles, with large numbers of admissions). Our second modification therefore scales (B.1) by $legal_{c,1988}$:

[1] a_{cst} could also be immigrants from country c admitted in year t, where country c is within world region s.

(B.2) $$\frac{a_{cst}}{legal_{c,1988}} = \delta_c + \gamma_t + \sum_{\tau>1988} \theta_\tau D_t^\tau \left(\frac{lpr_{c,IRCA}}{legal_{c,1988}}\right) + e_{cst},$$

where $e_{cst} \equiv \tilde{e}_{cst}\left(\frac{1}{legal_{c,1988}}\right)$ and $\delta_c \equiv \widetilde{\delta_c}\left(\frac{1}{legal_{c,1988}}\right)$. The year effects in this model, the γ_t, derive from dividing the second term in the parentheses of (B.1) by $legal_{c,1988}$, which transforms it to $\sum_{\tau>1988} \gamma_\tau D_t^\tau \left(\frac{legal_{c,1988}}{legal_{c,1988}}\right) = \sum_{\tau>1988} \gamma_\tau D_t^\tau = \gamma_t$; the last step is just notation. Thus the γ_t continue to represent the impact of pre-existing LPRs on new arrivals, and the θ_τ continue to capture the differential impacts of IRCA LPRs on new arrivals. Intuitively, the coefficients of interest ask whether deviations of new arrivals from prior trends correlate with the "intensity" of IRCA as a legalization shock.

C. Modification 3: State-by-year effects

Our third modification to the stylized framework accounts for the possibility that the intensity of IRCA as a legalization shock, or $\frac{ircalpr_c}{legal_{c,1988}}$, may correlate with other, unobserved state-by-time varying determinants of admissions, e_{cst}. The modification is to include a full set of state-by-year effects in (B.2):

(B.3) $$\frac{a_{cst}}{legal_{c,1980}} = \delta_c + \gamma_{st} + \sum_{\tau>1988} \theta_\tau D_t^\tau \left(\frac{lpr_{c,IRCA}}{legal_{c,1988}}\right) + e_{cst}$$

Returning to the derivation of (B.2), one can see that this modification allows the existing stock of legal immigrants in c, $legal_{c,1988}$, to have effects on admissions that are not only time-varying, as in (B.1), but state-by-time varying. In the country-level analysis, the γ_{st} represent world region-by-year fixed effects.

D. Modification 4: Using $legal_{c,1980}$ instead of $legal_{c,1988}$

Our next modification deals with a practical data challenge: we do not observe $legal_{c,1988}$ and use $legal_{c,1980}$ as a proxy. The year 1980 is last possible pre-IRCA year we can reliably measure the stock of legal residents: it is the last year that the U.S. maintained an alien registry, and it also coincides with a census year in which we can get an estimated count of citizens. One might instead attempt to impute a stock as of 1988 by adding up arriving new Green Card admissions between 1980 and 1988 (perhaps somewhat discounted for return migration). But notice that this can be construed as just another small group of potential sponsors that might confound our estimates.

To see this more formally, suppose this other group of potential sponsors who came between 1980 and IRCA's legalizations were denoted $legal_{c,1981-87}$, in other words, $legal_{c,1988} = legal_{c,1980} + legal_{c,1981-87}$. Substituting into (B.1) *and* allowing for separate coefficient vectors on each set of resulting interaction terms yields:

$$a_{cst} = \widetilde{\delta_c} + \sum_{\tau>1988}\left(\theta_\tau D_t^\tau lpr_{c,IRCA} + \gamma_\tau D_t^\tau legal_{c,1980} + \lambda_\tau D_t^\tau legal_{c,1981-1987}\right) + \tilde{e}_{cst},$$

Now, when we divide through by $legal_{c,1980}$ (instead of $legal_{c,1988}$) and make modifications 1 to 3, we have an "extra" term in the estimation equation:

$$\frac{a_{cst}}{legal_{c,1980}} = \delta_c + \gamma_{st} + \sum_{\tau>1988} \theta_\tau D_t^\tau \left(\frac{lpr_{c,IRCA}}{legal_{c,1980}}\right) + \sum_{\tau>1988} \lambda_\tau D_t^\tau \left(\frac{legal_{c,1981-1987}}{legal_{c,1980}}\right) + e_{cst}$$

However, letting $\varepsilon_{cst} \equiv \sum_{\tau>1988} \lambda_\tau D_t^\tau \left(\frac{legal_{c,1981-1987}}{legal_{c,1980}}\right) + e_{cst}$, one can obtain:

(B.4) $$\frac{a_{cst}}{legal_{c,1980}} = \delta_c + \gamma_{st} + \sum_{\tau>1988} \theta_\tau D_t^\tau \left(\frac{lpr_{c,IRCA}}{legal_{c,1980}}\right) + \varepsilon_{cst}.$$

Thus it becomes clear that dividing by the 1980 stock rather than the 1988 stock will only be an issue if the omitted terms $\sum_{\tau>1988} \lambda_\tau D_t^\tau \left(\frac{legal_{c,1981-1987}}{legal_{c,1980}}\right)$ are correlated with the legalization ratio, $\left(\frac{lpr_{c,IRCA}}{legal_{c,1980}}\right)$. Table A4 Panel B shows that there is not such a correlation for the cross-metro analysis, supporting the use of the 1980 proxy for our estimates; Table A7 Panel B presents comparable evidence for the cross-country analysis. In addition, Figures 4, 5, and A2 show no evidence of pre-trends in any admissions class (starting in 1983 due to data constraints), which is a sufficient condition for this result.

E. Modification 5: Additional interactions to test for pre-trends

Our final modification is to allow for such a test (for pre-trends) by expanding the model to include interactions between the legalization ratio and dummies for years prior to 1988:

(B.5) $$\frac{a_{cst}}{legal_{c,1980}} = \delta_c + \gamma_{st} + \sum_{\tau\neq1988} \theta_\tau D_t^\tau \left(\frac{lpr_{c,IRCA}}{legal_{c,1980}}\right) + \varepsilon_{cst}$$

This is the estimating equation in column 1 of Tables 2 and 4 for the cross-metro and cross-country analyses, respectively.

III. Comment on Alternative Scaling

Suppose that modification 2 had scaled by, for example, 1980 population, pop_c. The ultimate estimating equation would then have to be:

(B.6) $$\frac{a_{cst}}{pop_c} = \delta_c + \sum_{\tau>1988} \left(\theta_\tau D_t^\tau \frac{lpr_{c,IRCA}}{pop_c} + \gamma_\tau D_t^\tau \frac{legal_{c,1980}}{pop_c} + \lambda_\tau D_t^\tau \frac{legal_{c,1981-1987}}{pop_c}\right) + e_{cst}$$

Thus, to identify the coefficient vector of interest, θ_τ, would require multiple additional controls. Our preferred estimating equation is simpler and delivers the desired parameter estimates.

Appendix C: Measurement Error in the Scaling Factor

Both our treatment measure and our outcomes are scaled by a potentially mismeasured estimate of the number of legal residents prior to IRCA, $legal_{c,1980}$. Part of this scaling factor comes from the *Alien Address Reports*, which as a 100% sample of LPRs in 1980 should in principle be measured without error. However, we could only estimate the number of citizens using the 5% sample from the 1980 Census.

Below we show that, as a result of mismeasurement, our application could exhibit a variant of "division bias," a term coined by Borjas (1980). Because $legal_{c,1980}$ enters both sides and is in the denominator, measurement error is non-classical and does not necessarily lead OLS estimates to be biased toward zero. In fact, simulations using realistic levels of error suggest that random mistakes in $legal_{c,1980}$ if anything likely bias our OLS estimates slightly *upward* in magnitude. Moreover, simulations incorporating (non-random) miscounts for Mexicans (suggested in Warren and Passel (1987); Borjas, Freeman and Lang (1991)) suggest that the resulting bias in our OLS estimates is again small.

I. Baseline model

For expository ease, let y_{ct} represent Mexican LPR flows in MSA c in year t, x_c be Mexican IRCA LPRs in MSA c, and w_c be legal Mexicans in c in 1980 (i.e., $legal_{c,1980}$). In any given year, we are effectively interested in the simple bivariate relationship:

$$\frac{y_{ct}}{w_c} = \beta_{0t} + \beta_{1t}\frac{x_c}{w_c} + \varepsilon_{ct}. \tag{C.0}$$

If $\text{cov}\left(\varepsilon_{ct}, \frac{x_c}{w_c}\right) = 0$, the OLS estimator of β_{1t} is consistent:

$$\text{plim}\hat{\beta}_{1t}^{OLS1} = \text{plim}\frac{\widehat{\text{cov}}\left(\frac{y_{ct}}{w_c}, \frac{x_c}{w_c}\right)}{\widehat{\text{var}}\left(\frac{x_c}{w_c}\right)} = \beta_{1t} + \frac{\text{cov}\left(\varepsilon_{ct}, \frac{x_c}{w_c}\right)}{\text{var}\left(\frac{x_c}{w_c}\right)} = \beta_{1t}. \tag{C.1}$$

However, we instead observe a noisy measure of legal Mexicans in 1980, w_c. An assumption of classical measurement error is inappropriate, since the magnitude of the measurement error is arguably proportional to w_c, as was assumed in Borjas (1980); note in particular that sample variation in a count is proportional. For example, the distribution of the measurement error for Los Angeles, CA ($w_{LA} \approx 329{,}000$) is probably more spread than that for Oakland, CA ($w_{Oakland} \approx 23{,}000$).

Specifying this noisy measure as $\widehat{w}_c = w_c(1 + u_c)$, in practice we thus regress $\frac{y_{ct}}{w_c(1+u_c)}$ on $\frac{x_{ct}}{w_c(1+u_c)}$. What is the bias in OLS? Substituting for $\frac{y_{ct}}{w_c}$ using (C.0), we can express the OLS estimator of β_{1t} based on noisy data as:

$$\hat{\beta}_{1t}^{OLS2} = \frac{\widehat{\text{cov}}\left(\frac{y_{ct}}{w_c(1+u_c)}, \frac{x_c}{w_c(1+u_c)}\right)}{\widehat{\text{var}}\left(\frac{x_c}{w_c(1+u_c)}\right)} = \frac{\widehat{\text{cov}}\left(\frac{1}{1+u_c}\left(\beta_{0t} + \beta_{1t}\frac{x_c}{w_c} + \varepsilon_{ct}\right), \frac{x_c}{w_c(1+u_c)}\right)}{\widehat{\text{var}}\left(\frac{x_c}{w_c(1+u_c)}\right)}.$$

$\hat{\beta}_{1t}^{OLS2}$ does not converge to β_{1t}. Assuming, for example, that u_c is uncorrelated with both the signal, w_c, and the error term in (C.0), ε_{ct}, the probability limit of $\hat{\beta}_{1t}^{OLS2}$ is:

$$\text{(C.2)} \qquad \text{plim}\hat{\beta}_{1t}^{OLS2} = \beta_{1t} + \beta_{0t}\tau + \frac{\text{cov}\left(\frac{\varepsilon_{ct}}{1+u_c}, \frac{x_c}{w_c(1+u_c)}\right)}{\text{var}\left(\frac{x_c}{w_c(1+u_c)}\right)},$$

where τ is the slope parameter from a (hypothetical) regression of $\frac{w_c}{w_c(1+u_c)}$ on $\frac{x_c}{w_c(1+u_c)}$.[1] Even if $\text{cov}\left(\varepsilon_{ct}, \frac{x_c}{w_c}\right) = 0$, as assumed above, the third term of (C.2) is also not necessarily zero. The direction of the asymptotic bias of $\hat{\beta}_{1t}^{OLS2}$ therefore depends on the signs and relative magnitudes of more than one unknown parameter.

Our coefficients of interest are, however, *changes* in these slope estimates across years, e.g., from right before IRCA legalization to 10 years after, i.e.,

$$\hat{\theta}^{OLS2} = \hat{\beta}_{1POST}^{OLS2} - \hat{\beta}_{1PRE}^{OLS2}.$$

The asymptotic bias of this estimator is given by:

$$\text{(C.3)} \qquad (\beta_{0POST} - \beta_{0PRE})\tau + \left(\frac{\text{cov}\left(\frac{\varepsilon_{cPOST}}{1+u_c}, \frac{x_c}{w_c(1+u_c)}\right)}{\text{var}\left(\frac{x_c}{w_c(1+u_c)}\right)} - \frac{\text{cov}\left(\frac{\varepsilon_{cPRE}}{1+u_c}, \frac{x_c}{w_c(1+u_c)}\right)}{\text{var}\left(\frac{x_c}{w_c(1+u_c)}\right)} \right).$$

It is difficult to sign the second term of this expression without additional information. However, the first term is arguably positive. First, the β_{0t} represent year fixed effects, which in our model represent time-varying effects of w_c (i.e., $legal_{c,1980}$) on y_{ct} (i.e., LPR flows or a_{cst}). (See Online Appendix B.) We do not show estimates of these year fixed effects in the paper, but they are consistently positive in the post-IRCA period. Second, even though the regression is hypothetical, τ is arguably positive: MSAs with more legal Mexicans in 1980 likely had more IRCA applicants, due to the importance of enclaves in settlement patterns.

II. Instrumental Variables

In the standard classical measurement error/attenuation bias case, instrumenting for one noisy measure of a continuous treatment with another will generate a consistent estimate of the treatment effect. Such a result does not hold in this case. Even if we assume that u_c is uncorrelated with w_c and ε_{ct}, the proportional nature of the measurement error (or the sensitivity to scale) means that the measurement error is non-classical: $\widehat{w}_c = w_c(1 + u_c) = w_c + w_c u_c$, so the "signal," w_c, is correlated with the "noise," $w_c u_c$.

Returning to the bivariate slope, suppose that we regress $\frac{y_{ct}}{w_c(1+u_c)}$ on $\frac{x_c}{w_c(1+u_c)}$, but now instrument for $\frac{x_c}{w_c(1+u_c)}$ with $\frac{x_c}{w_c(1+\kappa_c)}$ (i.e., using another noisy measure of w_c) to construct an instrument. κ_c has the same properties as u_c: it is uncorrelated with both w_c and ε_{ct}. Furthermore assume that $\text{cov}(u_c, \kappa_c) = 0$. The two-stage least squares (2SLS) estimator of β_{1t} is thus:

[1] That is, $\tau = \text{cov}\left(\frac{w_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+u_c)}\right) / \text{var}\left(\frac{x_c}{w_c(1+u_c)}\right)$. Notice that $\tau = 0$ if $u_c = 0$ for all c.

$$\hat{\beta}_{1t}^{2SLS} = \frac{\widehat{\text{cov}}\left(\frac{y_{ct}}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)}{\widehat{\text{cov}}\left(\frac{x_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)}$$

Substituting with the true model (C.0) as above, we arrive at:

$$\hat{\beta}_{1t}^{2SLS} = \frac{\widehat{\text{cov}}\left(\frac{1}{1+u_c}\left(\beta_{0t}+\beta_{1t}\frac{x_c}{w_c}+\varepsilon_{ct}\right), \frac{x_c}{w_c(1+\kappa_c)}\right)}{\widehat{\text{cov}}\left(\frac{x_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)} \quad .$$

$\hat{\beta}_{1t}^{2SLS}$ then converges to:

$$\text{(C.4)} \qquad \text{plim}\hat{\beta}_{1t}^{2SLS} = \beta_{1t} + \beta_{0t}\frac{\text{cov}\left(\frac{w_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)}{\text{cov}\left(\frac{x_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)} + \frac{\text{cov}\left(\frac{\varepsilon_{ct}}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)}{\text{cov}\left(\frac{x_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right)}.$$

$\text{plim}\hat{\beta}_{1t}^{2SLS} = \beta_{1t}$ only if the second and third terms and equal and opposite in sign or if $\text{cov}\left(\frac{w_c}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right) = \text{cov}\left(\frac{\varepsilon_{ct}}{w_c(1+u_c)}, \frac{x_c}{w_c(1+\kappa_c)}\right) = 0$. There is no reason to expect either of these conditions to hold. 2SLS estimates of the parameter of interest, i.e., $\hat{\theta}^{2SLS} = \hat{\beta}_{1POST}^{2SLS} - \hat{\beta}_{1PRE}^{2SLS}$, are therefore likely inconsistent as well.

III. Simulations

These derivations imply that measurement error could be biasing our baseline OLS estimates, and that 2SLS will not eliminate this bias. The question then becomes: what direction is the bias in OLS, and is it big? As noted above, the year fixed effects for $t > 1988$ are consistently positive. Thus, empirical evidence suggests that $\beta_{0POST} - \beta_{0PRE} > 0$. However, this does not necessarily imply that the entire expression in (C.3) is positive or provide any insight into its magnitude.

We have approached signing the bias in two ways, both of which involve changing the amount of noise in the estimate of w_c, i.e., $legal_{c,1980}$. We discuss the findings for the metro area level analysis first.

Cross-Metro Analysis for Mexicans

First, we treat observed values of $legal_{c,1980}$ as being measured without error and add normally distributed proportional noise, u_c, where $\widehat{w}_c = w_c(1 + u_c)$. Throughout, we assume u_c is mean zero and uncorrelated with w_c and ε_{cst}. However, we allow $\text{var}(u_c)$ to vary and examine how estimates of θ change.

Table C1 columns 2 to 4 show the results of these simulations with increasing amounts of noise, while column 1 showing our original "noiseless" estimates for comparison (from Table 2, row 1, column 1). Column 4 allows for a great deal of measurement error: here, we assume that $\text{var}(u_c) = 0.2^2$, or that $\text{sd}(u_c) = 0.2$, which is close to the largest variance possible for a normally distributed u_c such that $\widehat{w}_c$ can remain positive. Columns 2 and 3 show estimates in which $\text{sd}(u_c)$ is smaller, 0.05 and 0.1, respectively. Across 1000 simulations under each of these assumptions, the average value of $\hat{\theta}$ for all family sponsored LPRs, shown in Panel A, is larger. Compared to the baseline estimate of 1.03, we obtain estimates of 1.035, 1.05, and 1.112,

respectively, with sd(u_c) =0.05, 0.1, and 0.2. These findings suggest that a noisy denominator in our treatment and outcome variables generates an *upward* bias in $\hat{\theta}$ (panel B) though not a substantial one in magnitude.

Second, we do a simulation based on estimated sample variation, the source of noise in our estimates of $legal_{c,1980}$. Note that $legal_{c,1980} = citizen_{c,1980} + lpr_{c,1980}$ where $citizen_{c,1980}$ is U.S. naturalized citizens from Mexico estimated in the 1980 Census (the source of sample variation) and $lpr_{c,1980}$ is the number of Mexican LPRs in 1980 from the *Alien Address Reports* (likely to be well measured). To figure out how the sample variation in estimates of $citizen_{c,1980}$ impact our estimates, we took 1000 random samples (with replacement) out of the 1980 Census data on Mexicans, stratified on MSA, and used them to construct 1000 alternative estimates of $citizen_{c,1980}$ for each MSA, and 1000 corresponding estimates of $legal_{c,1980} = citizen_{c,1980} + lpr_{c,1980}$ (where $lpr_{c,1980}$ stays the same).[2] We then ran the regressions with each of these alternative denominators. Note that this will tell us the direction and rough magnitude of bias from sample variation derived from samples of the size that are available in the Census.

Table C1 Column 5 shows the results. The average estimate with these 1000 alternative noisy denominators is larger – 1.067 versus 1.03 in our original sample. This suggests that sample variation in the denominator biases our slope estimate upward a bit, just as the simulated noise showed. But it again suggests that this bias is small.

Cross-Country Analysis

Random mismeasurement of the scaling factor could also bias the cross-country analysis. Our simulation of this, paralleling what we did in Table C1, is shown in Table C3. This time, the bias is essentially zero. Recall from the formulas above this is possible if the year effects are small (and they are).

IV. Under- and Overcounts and Other Systematic Mismeasurement

A more worrisome issue potentially would be more systematic errors in our estimate of $legal_{c,1980}$ that are correlated with the share unauthorized in 1980. This is a concern due to potential under- and over-counts of Mexicans in the 1980 Census of population.

As a reminder, $legal_{c,1980} = citizen_{c,1980} + lpr_{c,1980}$ where $citizen_{c,1980}$ is U.S. naturalized citizens from Mexico, estimated in the 1980 Census, and $lpr_{c,1980}$ are permanent residents from Mexico, measured from the registry. Researchers have raised the possibility that the 1980 Census may have systematically miscounted Mexicans. Warren and Passel (1987), for example, used historical naturalization statistics to claim that many more Mexicans in the 1980 Census claimed to be naturalized U.S. citizens than had actually naturalized; this may have occurred because some Mexicans wanted to hide their true legal status from enumerators. If so, we would systematically overstate $legal_{c,1980}$ in places with more unauthorized immigrants (many of whom would later go on to obtain status through IRCA). In contrast, Borjas, Freeman and Lang [BFL] (1991) use administrative data to argue the 1980 Census undercounted Mexicans by 25%.

[2] Below we will comment on error in $lpr_{c,1980}$.

To determine how such problems might bias our slope estimates, we added corrections for each to our data. In particular, taking Warren and Passel (1987), we created the alternative denominator $legal'_{c,1980} = citizen_{c,1980} + lpr_{c,1980} \times 205/580$, where $205/580$ is the ratio of the number of "true" Mexican-born U.S. citizens to the number that self-reported this status in the Census according to Warren and Passel (1980); this procedure thus deflates all self-reported citizen counts by this factor. For BFL, we constructed $legal''_{c,1980} = citizen_{c,1980} + pr_{c,1980}/0.75$ to reflect their finding that the 1980 Census undercounted the Mexican-born population by 25%.

Estimates with these alternative scalings are shown in Table C2, again along with our uncorrected estimate in column 1 (from Table 2). Note that, unlike Table C1, which provided "biased" estimates from adding noise, the alternative estimates in this table are the "corrected" ones that conceptually have *removed* the bias from systematic mismeasurement of $citizen_{c,1980}$.

The correction for Warren and Passel's estimated Census overcount is shown in column 2; the correction for BFL's undercount in column 3. If Warren and Passel's overcount scenario is correct, it implies that our estimate is biased downward relative to this corrected estimate by about 8% (1.03 vs 1.11 when corrected). If, in contrast, BFL's undercount scenario is correct, it implies that our estimate is biased upward by about 2% (1.03 vs 1.01 when corrected). We find these results reassuring. While they do suggest such miscounts could bias our estimates, for realistic values obtained from the literature, the bias appears to be small – citizens, after all, are a minority of legal Mexican residents.

As for systematic errors in the counts of legal residents in the 1980 Census for the cross-country analysis, the Census Bureau reports claims that the undercounts were generally quite low (based on "recatch" rates in a supplemental survey (e.g., U.S. Census Bureau, 1985)), and we have seen no alternative estimates of undercounts by country other than Mexico outside of Warren and Passel (1987). And while some non-Mexicans may have had a similar incentive to inflate their legal status as Mexicans appear to have done, Warren and Passel (1987) find much lower rates of naturalized citizen overcounts for origin countries other than Mexico.

So far, we haven't commented on measurement error in the other component, $lpr_{c,1980}$. This enters symmetrically into our analysis with $citizen_{c,1980}$, so mistakes in it will in principle have the same qualitative impact as mistakes in $citizen_{c,1980}$. For example, if it is somewhat underreported because some people fail to register, our estimates will be biased upward.[3] Random "noise" in the assignment of registered aliens to geographic areas (say, because of administrative errors in the database, or our ability to assign zip codes to metro areas) would also bias our estimates upward. In practice, however, $lpr_{c,1980}$ comes from an administrative register, reducing the chance for mistakes. Indeed, immigrants had a strong incentive to register to maintain their legal rights in the U.S., including the ability to sponsor relatives. Put differently,

[3] Though some sort of comparison to admissions data, Warren and Passel (1987) estimate that 7.3% of Mexican residents failed to register, though they provide no details of how that figure was arrived at. In addition, their "undercount" includes a broader set of residents which include not just those with Green Cards, the ones we are interested in, but also students and other temporary residents. While it seems plausible that underreporting in the registry for such temporary residents is non-zero, it is immaterial for our analysis.

any unregistered aliens would not have been allowed to sponsor relatives, and thus should not affect our outcome.

V. Summary

Summing up, measurement error in the scaling factor was a potentially important issue for our estimates. We have shown that the sign of any resulting bias in our estimates is not obvious. Unfortunately, there is no easy instrumental variables solution to this problem, either. However, for realistic scenarios in our data, we found that the bias is: (a) often upward, not downward; and (b) appears to be relatively small.

In retrospect, it makes sense that mistakes in the scaling factor are second order in practice. The major threat to valid identification is instead whether we can take the number legalized by IRCA as random, rather than what we scale it by. The balance test and lack of differential pre-trends in outcomes, and lack of any differential trends in admissions of non-relatives, suggest that we can.

VI. References

Borajs, George J. 1980. "The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias." *The Journal of Human Resources* 15(3): Summer, pp. 409-423.

Borjas, George J., Richard B. Freeman, Kevin Lang. 1991. "Undocumented Mexican-born Workers in the United States: How Many, How Permanent?" in John M. Abowd and Richard B. Freeman, eds, *Immigration, Trade and the Labor Market*. Chicago: University of Chicago Press.

U.S. Bureau of the Census. 1985. "The Coverage of Housing in the 1980 Census." Census of Population and Housing: 1980, Evaluation and Research Reports, PHC80-EI. Washington, D.C.: U.S. Government Printing Office.

Warren, Robert and Jeffery S. Passel. 1987. "A Count of the Uncountable: Estimates of the Undocumented Aliens Counted in the 1980 United States Census." *Demography* 24 (3): August, pp. 375-393.

Table C1. How Family Migration Response Changes When Scaled by Noisy Estimates of the Number of Mexicans Legal Residents in 1980

Original Estimate	Adding Proportional Noise in Number of Legal Mexicans with: sd = 0.05	sd = 0.1	sd = 0.2	Bootstrap Census Sample Variation
(1)	(2)	(3)	(4)	(5)
A. Mean Estimated Response of Family-Sponsored Migration with this Scaling Factor				
1.030	1.035	1.050	1.112	1.067
B. Implied Bias from This Noise:				
	0.005	0.021	0.083	0.037

Notes: In columns 2-5 mean estimates from 1,000 replications are shown. In columns 2-4, the number of Mexican legal residents was multiplied by a normally distributed random variable with a mean of 1 and a standard deviation shown in the column headers. In column 5, data on Mexicans in the 1980 Census of Population was randomly resampled, with replacement and stratified on metro area, and used to produce 1,000 alternative estimates of the number of Mexican citizens in each metro area in 1980 and added to Mexican permanent residents.

Table C2. Correcting Over- / Undercounts in the Census

Original Estimate	Correction for: Excess Mexican Naturalizations*	Mexican Under-count**
(1)	(2)	(3)
A. Mean Estimated Response		
1.030	1.112	1.011
B. Implied Bias in our Slope Estimate in Column (1):		
	-0.083	0.019

Notes: Estimates in column 2-3 scale by alternative estimates of the number of legally resident Mexicans in the U.S. in 1980 correcting for a *Overcount of naturalized Mexicans (Warren and Passel, 1987) in column 2 and a **25% undercount of Mexicans in column 3 (Borjas, Freeman and Lang, 1991).

Table C3. How Response of Family Migration Changes When Scaled by Noisy Estimates of the Number of Legal Residents in 1980: Cross-Country Analysis

Original Estimate	Adding Proportional Noise in Number of Legal Residents with: sd = 0.05	sd = 0.1	sd = 0.2	Bootstrap Census Sample Variation
(1)	(2)	(3)	(4)	(5)
A. Mean Estimated Response of Family-Sponsored Migration with this Scaling Factor				
1.737	1.740	1.741	1.772	1.736
B. Implied Bias from This Noise:				
	0.003	0.004	0.035	-0.001

Notes: In columns 2-5 mean estimates from 1,000 replications are shown. In columns 2-4, the number of legal residents was multiplied by a normally distributed random variable with a mean of 1 and a standard deviation shown in the column headers. In column 5, data on the foreign-born in the 1980 Census of Population was randomly resampled, with replacement and stratified on country, and used to produce 1,000 alternative estimates of the number of citizens from each country in 1980 and added to the number permanent residents from that country.