Measuring Sub-national Migration Networks using Matrículas Consulares∗

Maria Esther Caballero† Carnegie Mellon University
Brian C. Cadena‡ University of Colorado - Boulder
and IZA
Brian K. Kovak§ Carnegie Mellon University
NBER and IZA

June 2017

Abstract

Migrant networks affect the probability of migration, migrants’ location decisions, and their economic outcomes in destination communities. Yet available data sources on migrant networks either i) cover only a small fraction of sending- or receiving-country locations or ii) provide very coarse geographic information. In this paper, we show how to use administrative data from the Matrícula Consular de Alta Seguridad (MCAS) identification card program to measure migration networks with complete geographic coverage of both Mexico and the U.S. and detailed information on migrants’ sources and destinations. We first confirm the quality and representativeness of the MCAS data by comparing them with well-known household surveys in Mexico and the U.S., finding strong agreement on the migrant location distributions available across datasets. We then document substantial differences in the distribution of destinations among migrants from different municipios within the same source state, which implies that state-based network measures are at best a noisy measure of network connections based on place of birth. We conclude by demonstrating how these detailed migration network data can be used to study the effects of destination-specific conditions on migration patterns. We find that an Arizona law reducing employment opportunities for unauthorized migrants resulted in decreased emigration from and increased return migration to Mexican source regions with pre-existing strong network ties to that state.

Keywords: international migration, networks, immigration law, Mexico, United States

∗This project was supported by a grant from the Berkman Faculty Development Fund at Carnegie Mellon University. The authors would like to thank Alexandra Chouldechova, Terra McKinnish, and Lowell Taylor, and participants at the 2017 Population Association of America Annual Meeting and the University of Oxford Workshop on Immigration, Health, and Well-Being for helpful comments. Remaining errors are our own.

†Carnegie Mellon, Heinz College, 4800 Forbes Ave., Pittsburgh, PA 15213; mcaballe@andrew.cmu.edu
‡Univ. of Colorado, Dept. of Economics, 256 UCB, Boulder, CO 80309; brian.cadena@colorado.edu; 303-492-7908
§Carnegie Mellon, Heinz College, 4800 Forbes Ave., Pittsburgh, PA 15213; bkovak@cmu.edu; 412-268-5223
1 Introduction

A large literature documents the central role of migration networks in shaping the experiences of international migrants and the broader effects of international migration. Networks increase the probability of migrating internationally by lowering the costs of migration. They influence migrants’ location and job choices, with migrants tending to cluster both geographically and in particular occupations. By providing access to information, jobs, and other assistance, networks tend to improve migrants’ labor market outcomes in the destination. Migration networks have a wide variety of other effects, including altering the set of individuals choosing to migrate, affecting migrants’ approach to crossing the border, and increasing investment in origin communities.

The literature on Mexico-U.S. migration has revealed the importance of networks in spite of substantial limitations in typical datasets. No nationally representative survey records the migration experiences of friends or relatives. As a result, many network measures rely on the migration experiences of other individuals born in the same geographic location. However, this approach also faces substantial data limitations. Large-scale household surveys do not generally capture both origin and destination information at the sub-national level. For example, the U.S. Decennial Census and American Community Surveys allow for detailed measurement of where Mexican immigrants live within the United States, and the Mexican Census allows for similarly detailed measurement of the geographical distribution of origin communities for emigrants to the United States, but neither of the surveys records any sub-national geography for the other side of the international border. This prevents a joint characterization of place-to-place networks for Mexican sources and U.S. destinations.

1 A vast literature confirms the importance of networks in increasing rates of international migration. See, for example, Massey (1986); Massey and Espinosa (1997); Palloni, Massey, Ceballos, Espinosa and Spittel (2001); Winters, de Janvry and Sadoulet (2001); Curran and Rivero-Fuentes (2003); Fussell and Massey (2004); Liu (2013); Garip and Asad (2016).

2 See Bartel (1989); Jaeger (2000); Bauer, Epstein and Gang (2002); Diaz McConnell (2008); Patel and Vella (2013); and Lafontaine and Tessada (2014).

3 See Munshi (2003) and Mundra and Rios-Avila (2016).

4 McKenzie and Rapoport (2010) show that network access disproportionately increases migration probability for less educated individuals. Dolfin and Genicot (2010) examine the effects of family and community networks on migration with and without the assistance of smugglers. Woodruff and Zenteno (2007) show that migration networks increase microenterprise development in source locations.
In this paper, we overcome this limitation using data from the Matrícula Consular de Alta Seguridad (MCAS) program from 2006 to 2013, which allow us to measure the distribution of U.S. destinations for migrants born in each Mexican municipio (county)\(^5\). The MCAS identity card is available to Mexican citizens living in the United States, and an applicant must provide their municipio of birth in Mexico and their current address within the U.S. The Mexican government produces detailed tabulations of the universe of this administrative data at an annual frequency beginning in 2006. These cross-tabulations provide a complete characterization of the distribution of U.S. destinations for card holders born in each Mexican municipio, and data are available covering more than 7 million cards issued during our sample period.

Our analysis begins by confirming the quality and representativeness of this data source by showing that it coincides with well-known nationally representative surveys in the U.S. and Mexico. The share of MCAS applications in each U.S. state closely matches the share of Mexican-born residents living in that state according to the American Community Survey. Additionally, a Mexican municipio’s share of Mexican emigrants as measured in the Mexican Census closely matches the share of MCAS applicants who report being born in that municipio. Given the strong agreement on these important marginal distributions, we show that joint source-destination information in MCAS matches standard survey data as well. Specifically, the distribution of U.S. destination states for migrants from each Mexican state is similar in the MCAS and in the Encuesta Nacional de la Dinámica Demográfica (ENADID). However, it is clear that the relatively small sample size in the ENADID limits its ability to characterize the migration network, even at the aggregate state level. Together, these results confirm the quality of the migration network information generated from MCAS data, providing confidence that more detailed measures of the joint distribution of sending and receiving locations are also informative.

Such detailed geographic information on sources and destinations is essential to constructing accurate local migration network measures. We demonstrate this in the MCAS data by comparing destination distributions based on state of birth to distributions based on municipio of birth and

\(^5\)Massey, Rugh and Pren (2010) use the same data source for 2006 to describe more aggregate migration networks at the Mexican state level.
find that, as a general rule, state-based measures are quite different from municipio-level measures. In fact, the variation in destination networks for source municipios within the same Mexican state is larger than the variation in destination networks among the source states themselves. Thus, assigning all migrants their state’s average destination network introduces substantial measurement error into an analysis of the role of local networks in an individual’s migration experience.

We are aware of three existing survey datasets that measure both source and destination locations for migrants from Mexico to the United States, but each one has important limitations. The Mexican Migration Project (MMP) is the basis for much of the prior work on Mexico-U.S. migration networks, providing excellent detail regarding migration experiences for those surveyed. However, the MMP surveys only a small number of communities in Mexico and is representative only of those communities (Massey and Zenteno 2000), while the MCAS data provide complete geographic coverage of Mexico and the U.S.

As mentioned above, the ENADID is nationally representative, but it measures sources and destinations only at the state level, a fairly aggregate level of geography, which obscures much of the local variation in birthplace migration networks. The Encuesta sobre Migración en la Frontera Norte (EMIF) similarly provides state-level sources and destinations, but covers only individuals passing through particular border-crossing locations, and records intended destinations rather than realized destinations. Moreover, the ENADID and EMIF both have relatively small sample sizes, which prevent precise measurements of migration networks, even at the aggregate state level.

We conclude that for many research questions, migration network measures using administrative data from the MCAS program represent a substantial improvement over the measures typically used in the literature. Not only do these data provide complete geographic coverage of Mexico and the U.S., but the large number of observations allows for a full characterization of the joint distribution of sending and receiving locations at an exceptionally fine level of geography. With this data source, it is possible to calculate the share of observed migrants living in each of more than 70 U.S. destinations for each of more than 2,000 municipios of birth in Mexico. We anticipate that network

---

6 As of October 2015, the MMP had surveyed 154 communities whose combined populations accounted for 1.03 percent of the Mexican population in 2010.
measures based on this dataset will open the door to numerous additional lines of research into the importance of birthplace networks in the international migration experience.

As an initial example of the type of research that can be conducted using this superior network measure, we examine the effect of the Legal Arizona Workers Act (LAWA) on migration flows in to and out of source regions in Mexico. This law required employers throughout Arizona to submit an electronic request to confirm every prospective employee’s legal authorization to work in the U.S. The law thus reduced the attractiveness of Arizona as a destination for potential migrants without legal status. Using the MCAS data, we calculate the initial share of migrants from each source region (state or municipio) that selected Arizona as a destination prior to LAWA’s passage. We then use Mexican Census data to determine how changes in migration flows depended on the initial importance of Arizona in a source region’s migration network.

We find that source regions with stronger initial migration connections to Arizona experienced larger decreases in emigration to the U.S. and larger increases in return migration from the U.S. than sources with initially weaker connections to Arizona. Further, we demonstrate the value of the geographic detail in the MCAS data by comparing the robustness of state-level analysis to municipio-level analysis. Not only does the municipio-level analysis yield greater precision and less influence of high-leverage outliers, but it also allows us to control for any time-varying unobservable push or pull factors common to municipios within the same Mexican state.

This example validates our MCAS-based migration network measure in a variety of ways. First, the fact that we find an important role for the local network in driving responses to LAWA further reinforces our conclusion that the MCAS network data are informative. Second, point-to-point migration network information permits stronger conclusions regarding the migration impacts of state-level migration restrictions. Prior work has shown that these restrictions drive shifts in immigrant populations away from the relevant state (Bohn, Lofstrom and Raphael 2014) and decrease planned border crossings with the relevant state as the intended destination (Hoekstra and Orozco-Aleman 2017). By measuring the importance of Arizona in each specific source location’s migrant network, we can examine the effect of LAWA at the sending community level. The results reveal
that not only did migrants shift away from Arizona as a destination, but the policy also led to an overall decline in net migration to the U.S. from network-connected source regions. Third, by measuring migration networks for detailed sending regions, we can avoid the influence of outliers and improve the precision of our migration estimates.

The remainder of the paper is organized as follows. The next section provides background on the MCAS program and administrative data. Section 3 provides comparisons between the MCAS and multiple U.S. and Mexican data sources, showing that MCAS-based calculations closely match available geographical migrant distributions from high quality household surveys. Section 4 demonstrates that networks operate at a fine level of geography, including revealing the substantial differences between municipio-level and state-level networks. Section 5 presents the analysis of LAWA, showing that sources more connected to Arizona experienced larger decreases in net migration to the U.S. after the law’s passage. Section 6 concludes.

2 Matrícula Consular de Alta Seguridad (MCAS)

2.1 Background

The matrícula consular is a document issued by the Mexican government, providing its citizens abroad with a form of identification in their country of residence. In the U.S., the matrícula provides proof of citizenship, identity, and residence without conferring any immigration status on the cardholder. It is used primarily for returning to Mexico, for opening bank accounts, as identification to law enforcement officials, and, in some U.S. states and cities, to obtain a driver’s license and to access basic government services.

The Mexican government has been issuing matrículas since 1871, but in the 1990s they were transformed into wallet-sized laminated cards resembling a driver’s license. In 2002 a more secure version called the Matrícula Consular de Alta Seguridad (MCAS) was introduced (IME 2004), and additional security measures were added in 2006. Since the most recent security update, all MCAS issuances are recorded and verified through a centralized database, accessed by the issuing consulate
(National Immigration Law Center 2015). This database of MCAS distributions forms the basis for the dataset we use to measure Mexico-U.S. migration networks.

In order to obtain a card, the applicant must make an appointment and attend the Mexican consulate corresponding to their place of residence in the U.S. The applicant must provide proof of Mexican citizenship, identity, and residence in the relevant consular area, and they must not have a criminal record or be subject to judicial or administrative actions in the U.S. or Mexico (Secretaría de Relaciones Exteriores 2016). Cards are issued to all qualifying Mexican citizen applicants irrespective of age or immigration status, though it is generally assumed that the majority of MCAS holders are unauthorized immigrants who have limited access to other official forms of identification in the U.S (Massey et al. 2010). The card is valid for 5 years, and it can be renewed when it expires, when the cardholder moves to another consular jurisdiction, or if it is lost or damaged. MCAS are considered valid proof of identification by a wide variety of financial institutions and police departments in the U.S., and 12 states and the District of Columbia accept them as proof of ID to obtain a driver’s license (Mathema 2015).

2.2 Data

The recorded information from each approved MCAS application includes each card recipient’s municipio (similar to county) of birth in Mexico and the U.S. state and consular area of current residence. Consular areas refer to the geographic area of the U.S. within the jurisdiction of each Mexican consulate. The governmental Institute for Mexicans Abroad (Instituto para los Mexicanos en el Exterior, IME) uses this database to produce tabulations of the numbers of cards issued in each year. These tabulations include the count of cards issued for each birth municipio and U.S. location pair. By combining published tabulations based on U.S. state of residence with separate tabulations based on consular area of residence, one can generate counts of card issuances to individuals living in the 75 mutually exclusive and exhaustive destination areas in the U.S., shown in Appendix Figure A-1. With more than 2,000 Mexican municipios and 75 U.S. destinations represented, the card issuance data provide very detailed information on point-to-point migration patterns between
Mexico and the U.S.

A large number of MCAS are issued each year, and more than 7 million cards were issued during the 2006-2013 period. Nearly all of these issues are new applications rather than renewals. To assess the scale of these numbers, we calculate the stock of valid cards outstanding in each year from 2011 to 2013. Since MCAS are valid for 5 years, we can measure the stock of valid cards in a given year by summing the numbers of cards issued during the preceding 5 years. For example, all cards issued from 2006-2010 are still valid during at least a portion of 2011. Table 1 compares the number of valid cards in 2011-2013 to the estimated Mexican-born population of the U.S. (calculated from the ACS) and the Pew Research Center’s estimates of the unauthorized Mexican-born population of the U.S. (Passell and Cohn 2014, Gonzalez-Barrera 2015). The first two columns show a quite consistent 38 percent share of Mexican-born population holding a valid MCAS in each year. This share is similar to the 46 percent share reported in Suro and Escobar (2006), and the difference may reflect either a modest decrease in takeup between 2006 and 2011 or the fact that we observe the population of cards issued rather than a sample. If one assumes that only unauthorized immigrants hold matrículas, since they are most likely to benefit from access to identification, then the MCAS data cover 75 to 80 percent of the unauthorized Mexican immigrants living in the U.S.

Nevertheless, applying for a MCAS is voluntary, and the cards are distributed to a self-selected population. To understand the potential selection into takeup, we analyzed data from another Pew survey, which interviewed individuals applying for matrículas at various Mexican consulates in 2004-2005 (Suro 2005). Table 2 shows mean demographic and educational characteristics for this sample of matrícula applicants in comparison to all Mexican-born U.S. residents in the 2005 American Community Survey. Men, younger adults, and those with lower educational attainment are overrepresented among matrícula applicants in comparison to the overall Mexican-born popu-

---

7 Appendix Table A-1 provides an annual breakdown of the number of card issuances. In addition, officials at IME were kind enough to provide annual summary statistics on the share of card issuances reflecting new cards vs. renewals. In every year, fewer than 3 percent are renewals.

8 The number of cards issued prior to 2006 is not available, which is why we conduct this exercise beginning only in 2011.

9 Each of these columns reports the total counts of individuals of all ages.

10 The survey covered consulates in Los Angeles, New York, Chicago, Atlanta, Dallas, Raleigh, and Fresno.

---
lation. Additionally, matrícula applicants were more likely to have arrived in the late 1990s and early 2000s compared to the overall Mexican-born population. Together, these results are consistent with the idea that matrícula applicants are primarily drawn from the population of recently arrived unauthorized immigrants who are most likely to benefit from having access to an official identification card under the MCAS program.

3 MCAS Data Quality

The preceding results suggest that the MCAS data provide very good coverage of the recently-arrived unauthorized immigrant population. In this section, we examine the quality of the data in measuring migration networks between Mexico and the U.S. and contrast the MCAS data with other nationally representative data sources that one might use to measure migration networks.\footnote{We do not make comparisons between MCAS and the Mexican Migration Project (MMP) data because MMP communities are generally far smaller than and not necessarily representative of their municipios, so there is no comparable geographic unit across the two data sets.}

3.1 MCAS Data Match High Quality Survey Data

To validate MCAS data as a potential source of migration network information, we compare the migration patterns measured using MCAS to those in the largest and highest quality household surveys in the U.S. and Mexico. We begin by comparing estimates of the marginal distributions, i.e. source locations in Mexico and destination locations in the U.S.\footnote{Throughout this analysis, we use the U.S. state as the destination-level geography. Much of the analysis relies on the public-use microdata version of the ACS (Ruggles, Alexander, Genadek, Goeken, Schroeder and Sobek 2010), and the sub-state geographical definitions in that data source (PUMAs) do not align well with the consular areas. For consistency, therefore, we adopt the US state as the definition of a destination. The consular areas, however, are composed of US counties, and future work can match US datasets with county-level coverage to the destination geography shown in Figure A-1.}

Figure 1 provides the first such comparison, showing strong agreement between MCAS data and American Community Survey (ACS) data regarding the distribution of Mexicans across U.S. destination states. To construct this figure, we use the MCAS cards issued from 2006-2010, all of which were still valid on January 1, 2011. We then calculate the share of these cards reporting
a migrant’s residence in each of the 50 U.S. states plus the District of Columbia at the time of application. We construct a similar measure of the share of Mexican-born residents living in each state using the 2010 and 2011 American Community Surveys. Because the ACS is conducted throughout the year, by combining the 2010 and 2011 samples, we obtain a measure centered on January 1, 2011. We then compare the two distributions using a scatter plot, with the MCAS-based shares on the y-axis and the ACS-based shares on the x-axis. Since Mexican population is distributed unevenly across U.S. states, we plot the natural log of the state shares, allowing one to visually compare the two data sources for large and small states on the same figure. We include a 45-degree line, which shows how the two sets of shares would relate if the datasets agreed perfectly. It is readily apparent that the two datasets strongly agree, with only minor deviations from the 45-degree line. Moreover, the largest differences appear in states with very small numbers of Mexican immigrants. We have created similar figures for each of the time periods available in the data, centered on the beginning of 2012, 2013, and 2014, revealing similar agreement between the two data sources.

Figure 2 Panel (a) examines the distribution of Mexican source states for migrants to the U.S. We use the 2010 Mexican Census for this comparison, taking advantage of a question that asks respondents whether anyone currently or formerly living in the household migrated internationally between June 2005 and June 2010. Since we know the household’s location, we can calculate each Mexican state’s share of individuals observed leaving for the U.S. during this time period. We compare the log of this share with similar source-state shares from the MCAS data covering 2006-2010. Again, the close agreement across datasets is visually apparent from the figure, with only minor deviations from the 45-degree line.

Figure 2 Panel (b) repeats the analysis of Panel (a) but changes the geographic definition to the Mexican municipio. Although these measures are somewhat noisier, especially in municipios with smaller populations, the agreement between MCAS and the Mexican Census at this fine level of geographic detail is remarkable and reflects the very large number of migrants present in the MCAS.
data, which facilitates high quality measures of migration patterns even for small geographic areas. These results show that the MCAS data closely coincide with the best available measures of source and destination information for Mexican migrants to the U.S.

### 3.2 ENADID and EMIF as alternatives

We next consider whether either of the other two datasets with complete geographic coverage of sending and receiving locations compares as well with these Census distributions. We begin with the *Encuesta Nacional de la Dinámica Demográfica* (ENADID), which reports each migrant’s state of residence in the U.S. in addition to their source location. The question identifying migrants is similar to the one in the Mexican Census, and it allows us to observe migrants who left for the U.S. between August 2009 and August 2014. Figure 3 Panel (a) is constructed analogously to Figure 1, comparing the distribution of destination states among migrants in the ENADID who left during 2009-2014 to the 2014 ACS. The ENADID destination measure performs reasonably well, but there are important ways in which the MCAS data in Figure 1 align more closely with the baseline ACS distribution. First, Figure 3 Panel (a) includes only 41 U.S. states, as there are ten U.S. states that were not reported as the destination for any migrants observed in the ENADID. Second, the figure shows that the distribution aligns closely for large-population states such as Texas and California, but there is noticeably more disagreement between the two data sources for mid-range population states (those with log shares between -4 and -6). The MCAS data matched very closely in this range; larger differences from the ACS were observed only for very small states that the ENADID omits entirely. These differences highlight the primary shortcoming of the ENADID: a much smaller sample of migrants than in the MCAS database.

Figure 3 Panel (b) provides a comparison of source states analogous to Figure 2 Panel (a), using the ENADID in place of the MCAS. Again, the ENADID performs reasonably well, although the share comparisons are not as tightly clustered around the 45-degree line as are the observations.

---

13Recall that the MCAS data provide information on the universe of card issuances, while the Census migration question is asked only of a 10 percent subsample. If the total number of observations were the only relevant criterion, the MCAS data would be superior even to the Census data.
using the MCAS data, likely resulting from the ENADID’s smaller sample size. Appendix Figures A-2 and A-3 present analogous comparisons using the EMIF. The EMIF performs no better than the ENADID data, and the EMIF destination distribution is much less consistent with the ACS. Because the ENADID appears to be the best survey option, we focus our remaining comparisons on the MCAS and ENADID data.

3.3 Comparison of Networks in MCAS and ENADID

We next turn to a comparison of the geographic migration network measures available in both the MCAS and ENADID. Our analysis focuses on each dataset’s measure of where migrants from a given Mexican state are likely to locate within the United States. Using MCAS applications from 2009-2013 and ENADID data from 2014, we calculate for each source state the share of migrants selecting each U.S. state as their place of residence. Figure 4 graphs the relationship between the natural log of these shares for each source-destination combination, labeling each observation with the U.S. destination state. Although there is some agreement between these two data sources, the influence of the ENADID’s smaller sample size is readily apparent. Notably, there are relatively few observations in the lower left quadrant of the figure. These “missing” observations reflect destinations that are relatively uncommon in the ENADID or that fail to appear at all due to its small sample size.

In fact, the sample sizes in the ENADID are sufficiently small that empty cells are guaranteed to occur. Mexican states with fewer than 51 observed out-migrants must have at least some empty source-destination cells. As shown in detail in Appendix Figure A-4, nearly 80 percent of Mexican source states have fewer than 51 observed migrants in the ENADID. In contrast, every Mexican state has at least 51 individual MCAS applicants observed over the five-year period, and the vast majority of states have more than 100,000 observed migrants. Even at the relatively aggregate state level, the ENADID simply does not observe a sufficient number of migrants to credibly estimate source-specific destination distributions for the majority of Mexican sources.

This discrepancy likely occurs because the EMIF asks about a migrant’s intended destination, which is subject to change.
Taken together, the results in this section imply that the MCAS data provide an excellent way of measuring place-based migration networks between Mexico and the U.S. The distributions of both sources and destinations closely match the highest quality available survey datasets, and the joint distribution of sources and destinations corresponds reasonably well with the distribution observed in the ENADID, despite the latter data source’s small sample size. Finally, the number of observations per cell is orders of magnitude larger in the MCAS data, which both increases the precision of the estimated share of migrants from a source choosing a particular destination and also greatly reduces the potential for entirely missing source-destination pairs with small numbers of migrants.

4 Networks Operate at Fine Levels of Geography

Having established that MCAS data are a superior resource for examining Mexico-U.S. migration networks, we next examine the importance of using networks based on fine rather than coarse geography. Our primary analytical tool is a measure of the dissimilarity of two discrete distributions, as described by Duncan (1957). This index quantifies the difference in the destination distribution between individuals from source $s$ and individuals from a reference group.

$$\Delta_s \equiv \frac{1}{2} \sum_{d=1}^{D} \left| \pi_{sd} - \pi_{d}^{\text{ref}} \right|,$$

where $\pi_{sd} \equiv \frac{N_{sd}}{N_s}$ is the share of emigrants from source $s$ residing in destination $d$, and $\pi_{d}^{\text{ref}}$ is the share of the reference population in destination $d$. This index is bounded between zero and one, with zero representing identical distributions and one representing distributions with no overlap whatsoever. The magnitude of the measure can be interpreted as the share of migrants from source $s$ who would need to be reallocated across destinations in order to exactly match the destination distribution of the reference group. Note that Figures 1 through 3 report this dissimilarity measure for the comparisons reflected in each figure.

We begin by using the MCAS data from 2006-2010 to document the variation in network con-
nnections based on migrants’ state of birth. Specifically, we examine how each Mexican state’s distribution of migrant destinations in the U.S. compares to the destination distribution of all Mexican emigrants. In this case, $s$ refers to Mexican states, $d$ refers to U.S. destination states, and $\text{ref}$ refers to all Mexican migrants in the U.S. In order to clarify how this measure is calculated, Figure 5 shows the difference between the destination shares for migrants from Michoacán and the destination share for all Mexican migrants ($\pi_{sd} - \pi_{d}^\text{ref}$).\(^{15}\) Compared to the average Mexican migrant, migrants from Michoacán are more likely to live on the West Coast and in Illinois, and they are much less likely to live in Texas. Adding up the absolute value of these differences yields a value of $\Delta_s = 0.21$, indicating that 21 percent of the migrants from Michoacán would need to relocate within the U.S. in order to match the overall distribution of Mexican migrants’ chosen destinations. We repeat this analysis for each sending state, and Figure 6 provides a histogram showing the distribution of $\Delta_s$ for all Mexican source states. Most states’ measures are reasonably high, meaning that different Mexican source states send migrants to quite different sets of locations in the U.S.

Figure 7 Panel (a) provides the distribution of the same dissimilarity measure calculated at the municipio rather than state level, and it reveals even more variation. Although a few municipios have destination distributions close to the national average (small dissimilarity measures), most are quite different (large dissimilarity measures). Many source municipios would require more than half of their migrants to choose different destinations in order to match the destination distribution of all Mexican migrants.

We next examine variation in network connections among municipios in the same sending state. For each municipio, we calculate a new version of $\Delta_s$ using the Mexican state containing the source municipio as the reference group. Figure 7 Panel (b) provides the distribution of these within-state dissimilarity measures.\(^{16}\) This distribution is noticeably shifted to left compared to Panel (a), which confirms that the destinations selected by migrants from a given municipio are, in general,

\(^{15}\)The destination distribution for migrants from Michoacán ($\pi_{sd}$) appears in Appendix Figure A-5. Appendix Figure A-6 shows the overall distribution of migrant destinations, which is used for comparison ($\pi_{d}^\text{ref}$).

\(^{16}\)For more detail on the data points underlying this histogram, see Appendix A.5 which includes example municipio distributions and the relevant comparisons with those municipios’ state.
more similar to their state’s distribution than they are to the national average. Yet the histogram reveals that it is very common for two municipios within the same state to have very different destination distributions. Thus, for many municipios, the state distribution is a poor proxy for the true municipio-level distribution.

As an example, consider two municipios in the state of Michoacán: Ciudad Hidalgo and Tiquicheo. These source locations are only a 3 hour drive apart, but their destination distributions differ sharply. Migrants from Hidalgo settle primarily in Illinois (likely Chicago), while more than two thirds of emigrants from Tiquicheo reside in Texas. This difference in destinations occurs within the same source state, ruling out the possibility that it arises due to other factors affecting destination choice such as distance, climate similarity, etc. It is also noteworthy that neither distribution is particularly close to the state level distribution, which is more concentrated in California.

The results in this section have important implications for researchers using networks as a source of identifying variation. First, state of birth network measures are a poor proxy for networks that operate at a finer level of geography. Second, although source-state-level migration networks differ from one another, there is even more variation in the destination distribution across different municipios located in the same state. The ability to construct a network measure based on all individuals municipio of birth, therefore, makes the MCAS data a particularly valuable resource for researchers studying Mexico-U.S. migration networks.

5 Effects of LAWA on Migration in Network-connected Mexican Sending Regions

As an example of the value of this level of geographic detail, the following section examines the effect of the Legal Arizona Workers Act (LAWA) on migration rates into and out of sending regions in

\[\text{For the full distribution of destinations chosen by migrants from these two sources, see Appendix Figures A-7 and A-8.}\]
Mexico. This act, passed by the Arizona state legislature in 2007 with an effective date of January 2008, led to a decline in the “likely unauthorized” population living in Arizona (Bohn, Lofstrom and Raphael 2011). Data limitations, however, prevented an analysis of whether this decline reflected changes in international migration or simply the movement of unauthorized immigrants into different U.S. states. Our analysis leverages the MCAS data to show that emigration fell and return migration rose in source locations that were more exposed to LAWA through the migrant network. Thus, the decline in destination labor market opportunity driven by LAWA reduced the number of Mexican immigrants living in the United States from regions with ties to Arizona.

5.1 The Legal Arizona Workers Act

LAWA mandates the use of E-verify, an online system administered by the federal government, to verify prospective employees’ identity and authorization to work in the United States. The E-verify system compares Social Security numbers and names of new workers against a centralized database from the Social Security Administration (SSA) and the Department of Homeland Security (DHS). When there is no match between the employee’s name or social security number and the official records, the system sends a report of non-confirmation to the employer. The law imposes sanctions on employers who hire unauthorized workers, ranging from business license suspensions for the first offense to license revocation (Bohn et al. 2011).

LAWA’s main purpose was to increase the costs for employers hiring unauthorized migrants and for unauthorized employees looking for jobs. As a result, Arizona became a less desirable destination to live and work for immigrants without legal status. We examine the effect of LAWA on international migration by studying changes in migration patterns in Mexican sending regions connected to Arizona through the migrant network. Specifically, we assess whether Mexican sending regions initially more connected to Arizona experienced larger increases in return migration and larger decreases in emigration rates after LAWA was passed. Importantly, these analyses use Mexican Census data to measure migration rates in and out of Mexico; the MCAS data are used only to characterize the degree to which a sending location was initially connected to Arizona.
5.2 Effects of LAWA on Migration Rates

Our analysis treats the implementation of LAWA as a quasi-experiment that manipulates Mexican-born individuals’ U.S. job prospects. We expect that LAWA negatively affects the job prospects for all Mexican-born individuals without U.S. work authorization, but that these effects will be larger for individuals with strong network connections to Arizona. Because we cannot observe an individual’s network contacts directly, we rely on the geography-based networks observable in MCAS data. Specifically, we assume that LAWA’s effect on the labor market prospects in the U.S. for a migrant from region \( s \) is proportional to the share of MCAS card recipients from the same sending region who lived in Arizona in 2006.

We use the following specification to relate changes in migration rates from 2005 to 2010 to the migrant network:

\[
\Delta Y_s = \beta_0 + \beta_1 \pi_{s,2006} + \epsilon_s
\]

where \( \Delta Y_s = \ln(y_{s,2010}) - \ln(y_{s,2005}) \) is the change in the log of the return migration or emigration rate from 2005 to 2010 in Mexican sending region \( s \). We calculate migration rates using the 2010 Mexican Census and the 2005 Conteo de Población y Vivienda. \( \pi_{s,2006} \) is the share of emigrants from source \( s \) selecting Arizona as a residence in 2006, before LAWA was passed, using the MCAS data. \( \epsilon_s \) is an error term. \( \hat{\beta}_1 \) therefore captures the differential change in migration rates for Mexican source locations that were more connected to Arizona prior to LAWA. We conduct this analysis using states or municipios as source regions, treating states as independent observations and computing standard errors clustered at the state level when using municipios as the unit of analysis.

5.2.1 Return Migration Rates

The first specification uses return migration rates as the outcome variable. Return migrants are defined as individuals living in Mexico during the 2010 Census or 2005 Conteo reference period, but who lived in the U.S. five years before. The return migration rate is then the number of return
migrants divided by the source’s population at the beginning of each period (2000 or 2005).\footnote{Return migration flows are identified in the 2005 \textit{Conteo} and in the 2010 Mexican Census through a question that records country of residency five years prior to the date when the survey was administered. The count of return migrants does not include any individuals who were living in Mexico five years previously, but who moved to the U.S. and back within the five year window.}

Figure 8 Panel (a) presents the underlying data and the fitted values for equation (2) using the change in the natural log of this measure from 2005 to 2010 as the dependent variable. Consistent with expectations, return migration rates rose more in sending states with stronger network connections to Mexico. The first column of Panel A of Table 3 provides the coefficient estimates from this specification. To understand the magnitude of this effect, it is important to bear in mind that this time period saw dramatic declines in net migration to the United States overall, both through increased return migration and through decreased emigration. The average Mexican state’s return migration rate nearly quadrupled over this time period (rates were 0.3% on average in 2005 and 1.13 % on average in 2010). Yet these results imply that Mexican states with strong connections to Arizona saw larger increases in return migration than in states with weaker connections. Comparing a state like Sonora, where roughly 50 percent of migrants had historically settled in Arizona, to a state with no ties to Arizona, the growth in the return migration rate was roughly 30 percent larger than in the state with no connection.\footnote{The percentage difference in the growth of return migration rates would be $100 \times e^{0.558} - 1 = 32.1\%$. Compared to growth rates of roughly 300 percent in untreated states, this treated state would see a growth rate of roughly 400 percent.}

Columns (2) and (3) of Table 3 examine the robustness of this result. The second column in Panel A shows results when the observations are weighted by the 2000 Mexican population.\footnote{This weighting addresses the fact that population growth rates are heteroskedastic, with smaller populations experiencing more variable percentage growth in migration rates.} The third column provides results from a robust regression technique that reduces the impact of high leverage outliers. The point estimates are positive and quite similar across all three specifications, but the large standard error in column (3) makes clear that these state-level results are highly dependent on the inclusion of a single high-leverage observation (Sonora).

We now take advantage of the finer geographical detail available in the MCAS to conduct similar analysis using the Mexican \textit{municipio} as the unit of analysis. Panel B of Table 3 presents...
similar estimates for the change in log return migration rates from 2005 to 2010 at the Mexican 
municipio level. The first column provides the baseline estimates using municipios as observations, 
and Figure 8 Panel (b) provides the raw data and fitted line from this regression. The point 
estimate is comparable in magnitude with the first column of Panel A. Notably, the standard error 
is substantially smaller, and the scatter plot makes clear that there is no particular high leverage 
outlier driving the results. In fact, implementing the same robust regression technique used in the 
state-level analysis (column 3) leads to standard errors that are not much different from those in 
the baseline results in column (1).

An additional advantage of conducting a municipio-level analysis is the ability to add controls 
for other changes over time that affect return migration rates. Changing conditions in sending 
regions will change individuals’ incentives to return or to leave for the U.S. To the extent that these 
changes in conditions are correlated with a location’s ties to Arizona, they represent a threat to the 
causal interpretation of these regression results. In columns (4)-(6) we therefore add Mexican state 
fixed effects as additional controls. Because the dependent variable is already expressed as a long 
difference within municipios, these fixed effects remove the influence of any changes in the sending 
areas that affected migration rates similarly for all municipios within a Mexican state. For example, 
this specification controls for changes in state-level labor market conditions and changes in state-
level crime. This specification leverages the within-state variation in destinations shown in Figure 
7b and compares municipios that are geographically close to each other and yet are differentially 
connected to Arizona’s labor market. In each case, the results in columns (4)-(6) are similar to the 
corresponding results without fixed effects in columns (1)-(3). The ability to conduct analysis at the 
municipio level thus strengthens the results by increasing precision, by decreasing the importance 
of outliers, and by allowing for flexible controls for unobserved changes that could be correlated 
with the strength of a location’s ties to Arizona.
5.2.2 Emigration Rates

Having shown that LAWA increased the rate at which individuals return to Mexico, we now turn to the other component of net migration – emigration to the United States. Emigration is measured using the 2010 Census, which reports the year in which household members traveled to the U.S. We calculate the emigration rate as the number of people who reported emigrating in a given year divided by the source-area population in that year, for 2005 and 2010.

Table 4 is analogous to Table 3 and examines changes in the log of the emigration rate from 2005 to 2010 at the Mexican state and municipio levels. The scatter plot and fitted line for the regression in the first column of Panel A is provided in Figure 9 Panel (a); Figure 9 Panel (b) provides a similar graph for the specification in the first column of Panel B. On the whole, the results for emigration mirror the results for return migration, with sources more connected to Arizona seeing larger decreases in emigration from 2005 to 2010. Again, the importance of a single observation (Sonora) in the state-level results is apparent both visually in the scatter plot and in the large standard error in the third column of Panel A. In contrast, the municipio-level results are robust to mitigating the influence of high leverage outliers (columns 3 and 6 of Panel B). Further, the results are robust to the addition of state fixed effects (columns 4-6 of Panel B), and are in fact stronger when including these controls.

Together, these results demonstrate that the LAWA-induced declines in Arizona’s likely-unauthorized population documented by Bohn et al. (2014) occurred partly through substantial changes in international migration patterns. Return migration rose more sharply in Mexican source locations where migrants had historically settled in Arizona, and emigration fell more in those same locations. These effects of LAWA are, to our knowledge, a novel result, and they imply that policies affecting migrants’ job opportunities can be effective at deterring or reversing the flow of unauthorized migrants into the United States. Moreover, these results further strengthen our conclusion

---

21 A small number of emigrants who left in 2005 are not observable because the survey asks only about the start date of the most recent trip, so individuals who emigrated first in 2005, then returned to Mexico and then emigrated again are treated as having emigrated in the year of their most recent trip to the U.S.

22 Hoekstra and Orozco-Aleman (2017) examine a related question using a later Arizona law, SB 1070, which imposed unprecedented immigration enforcement measures. They use EMIF data to document decreased intended
that the migrant network measure based on MCAS data is informative. Had the data been sharply non-representative or highly noisy, we would not observe these important differences across Mexican source regions with strong initial connections to Arizona.

6 Conclusion

In this paper, we evaluate the use of administrative data from the Matrícula Consular de Alta Seguridad (MCAS) program to measure geographic migration networks between Mexico and the U.S. Unlike other available datasets that one could use to characterize these networks, the MCAS data provide very large sample sizes, detailed geographic identifiers, and complete geographic coverage of both Mexico and the U.S. Because the MCAS data do not comprise a traditional random sample drawn from a pre-specified population, however, it was necessary to determine whether the migration distributions calculated from MCAS are representative of migration networks for the overall population of Mexican-born individuals living in the U.S.

We established this representativeness by comparing the distributions of Mexican source states and U.S. destination states measured using MCAS and using each country’s Census of Population. The measures agree very closely across datasets. Alternative sources of network information such as the Encuesta Nacional de la Dinámica Demográfica (ENADID) and the Encuesta sobre Migración en la Frontera Norte (EMIF) either have sample sizes that are too small to capture enough migrants to accurately estimate the migrant network, or simply disagree with more credible Census measures. Moreover, the MCAS data contain far more detailed geographic information on Mexican source locations than any of these other sources of migration network information. We demonstrated that detailed sending regions within the same Mexican state regularly send migrants to very different sets of locations in the U.S. In fact, the amount of variation in destination distributions across municipios within the same Mexican state is substantially larger than the amount of variation across Mexican states. Using more aggregate state-level migration network information would therefore migration to Arizona among unauthorized Mexican migrants after the law’s passage. Our novel contribution is to document changes in realized return migration and emigration at the Mexican source level. Given the shortcomings of EMIF described in Section 3, such an analysis would not be feasible using that data source.
obscure these differences between detailed sending regions, likely reducing the apparent influence of migration networks on a variety of outcomes.

We reinforce these conclusions by using the detailed migration network data to study the effects of the 2008 Legal Arizona Workers Act on international migration between Mexico and the U.S. We find increased return migration to and decreased emigration from Mexican regions with stronger pre-existing migration ties to Arizona. These findings indicate that labor market interventions can strongly affect international migration patterns. Moreover, they confirm the empirical utility of our geographically detailed migration network measures. Analyses using state-level rather than municipio-level network information yield statistically weaker and less robust results.

Together, these findings imply that the MCAS data provide a valuable resource for researchers seeking to understand the influence of migration networks on migration behavior and on migrants’ social and economic outcomes.
References


The figure plots the distribution of Mexican-born individuals across U.S. destinations. Each point represents the natural log of the share of individuals in each dataset living in each U.S. state. The ACS sample includes Mexican-born individuals sampled in 2010 and 2011. The MCAS data include the universe of identity cards issued during 2006-2010. These cards were valid through the 2010-2011 time frame covered by the ACS sample. The 45-degree line, which would indicate perfect agreement between the two data sources, is shown for reference. The figure indicates a strong agreement between the two datasets, with only minor deviations from the 45 degree line. The largest differences are in U.S. states with historically low shares of Mexican migrants. The dissimilarity index, shown in the lower left corner, is defined in (1) and is interpreted as the share of individuals that would need to be reallocated to make the two datasets’ distributions match exactly.
The figures plot the distribution of Mexican source states and source municipios for migrants to the U.S. Each point represents the natural log of the share of individuals in each dataset from each Mexican state or municipio. The Mexican Census sample includes individuals who moved to the U.S. during the five year period from June 2005 to June 2010. The MCAS sample includes the universe of identity cards issued during 2006-2010. Vertical striping in Panel (b) reflects municipios with very small numbers of observations. The 45-degree line, which would indicate perfect agreement between the two data sources, is shown for reference. The figure indicates close agreement between the two datasets at the state level, but with more noise at the municipio level. In general, the biggest differences between the datasets appear in states with historically low shares of emigrants to the U.S. The dissimilarity index, shown in the lower left corner of each panel, is defined in (1) and is interpreted as the share of individuals that would need to be reallocated to make the two datasets’ distributions match exactly.
Figure 3: Comparisons of ENADID Source and Destination Distributions

(a) Distribution of U.S. Destination States: ENADID vs. ACS

Panel (a) plots the distribution of Mexicans across U.S. destination states. Each point represents the natural log of the share of individuals in each dataset living in each U.S. state. The ACS sample includes Mexican-born individuals arriving before or on 2014. The ENADID sample includes those who moved to the U.S. during the five year period from May 2009 through May 2014. The 45-degree line, which would indicate perfect agreement between the two data sources, is shown for reference. Both figures indicate that the MCAS data more closely matches the standard comparison datasets than does the ENADID. The dissimilarity index, shown in the lower left corner of each panel, is defined in (1) and is interpreted as the share of individuals that would need to be reallocated to make the two datasets’ distributions match exactly.

(b) Distribution of Mexican Source States: ENADID vs. Mexican Census

Panel (b) plots the distribution of Mexican source states for migrants to the U.S. Each point represents the natural log of the share of individuals in each dataset from each Mexican state. The Mexican Census sample includes individuals who moved to the U.S. during the five year period from June 2005 through June 2010. The ENADID sample includes those who moved to the U.S. during the five year period from May 2004 through May 2009. The dissimilarity index, shown in the lower left corner of each panel, is defined in (1) and is interpreted as the share of individuals that would need to be reallocated to make the two datasets’ distributions match exactly.
The figure plots the distribution U.S. destination states for migrants from each Mexican source state. Each point represents the natural log of the share of migrants living in the labeled U.S. state for each Mexican source state, from each dataset. MCAS data include the universe of identity cards issued during 2009-2013. The ENADID sample includes individuals who moved to the U.S. in the five year period from May 2009 through May 2014. The 45-degree line, which would indicate perfect agreement between the two data sources, is shown for reference. The figure reveals some agreement for U.S. states with large shares of Mexican migrants (e.g. Texas and California), but substantial disagreement for many smaller source-destination pairs. The small sample issue in ENADID is reflected in the lack of observations for source-destination pairs with small numbers of migrants. The asymmetry around the 45-degree line results from logging the shares; they sum to 1 across destinations for each source.
Figure 5: Difference in Destination Distribution for Migrants Born in Michoacán vs. All Mexican-Born

This figure shows the difference in the distributions of U.S. destination states for migrants born in Michoacán vs. all Mexican-born migrants, calculated using the universe of MCAS identity cards issued from 2006-2010. The distributions for each group are mapped separately in Appendix Figures A-6 and A-5. Darker blue (red) indicates that the state accounts for a larger (smaller) share of migrants from Michoacán than for all Mexican-born migrants. Migrants born in Michoacán are more likely to live in the West Coast and Illinois and less likely to live in Texas, relative to the average Mexican migrant.
This figure shows the distribution of dissimilarity indexes for the 32 Mexican states, using the universe of MCAS identity cards issued from 2006-2010. For each Mexican source state, we calculate \[ \frac{|D_{ij} - D_{ij}^{total}|}{D_{ij}^{total}} \], comparing the Mexican state’s distribution of migrants across U.S. destination states to the destination distribution for all Mexican-born migrants to the U.S. The resulting value (on the x-axis) is the share of individuals from the reference state that would need to be reallocated across destinations to match the overall distribution of Mexican-born migrants. The figure reveals that most Mexican states would need to reallocate more than 20 percent of their migrants across U.S. destinations to match the destination distribution of all Mexican migrants.
This figure shows the distribution of dissimilarity indexes for the 2,442 Mexican municipios, using the universe of MCAS identity cards issued from 2006-2010. In both panels, we calculate the U.S. destination state distribution for a given municipio to a reference distribution. In Panel (a), the reference distribution is the destination distribution for all Mexican migrants. In Panel (b) the reference distribution is the destination distribution for migrants from the Mexican state containing the relevant municipio. Panel (a) shows that many Mexican municipios would need to reallocate more than 50 percent of their migrants across U.S. destinations to match the destination distribution of all Mexican migrants. Panel (b) shows that many Mexican municipios would need to reallocate more than 30 percent of their migrants across U.S. destinations to match the destination distribution of their states, indicating quite substantial variation in migrant networks for municipios within the same Mexican state.
The figures plot the relationship between the change in log of return migration rate from 2005 to 2010 in Mexican sending regions ($\Delta Y_s$) vs. Arizona’s initial share of emigrants from each Mexican sending region ($\pi_s$) in 2006, before LAWA went into effect. Migration rates were calculated using the 2005 Mexican Conteo de Población y Vivienda and the 2010 Mexican Census, whereas $\pi_s$ was calculated using the 2006 MCAS. $\pi_s$ represents the exposure of Mexican sending regions to LAWA, so the positive relationship implies increased return migration to Mexican sources whose migrant networks were more exposed to Arizona. Panel (a) shows the relationship at the Mexican state level, and Panel (b) shows the relationship at the municiplio level. The regression lines are based on the regression model in (2), with no controls. See Table 3 for alternative specifications.
The figure plots the relationship between the change in log of emigration rate from 2005 to 2010 in Mexican sending regions ($\Delta Y_s$) vs. Arizona’s initial share of emigrants from each Mexican sending region ($\pi_s$) in 2006, before LAWA went into effect. Migration rates were calculated using the 2005 Mexican Conteo de Población y Vivienda and the 2010 Mexican Census, whereas $\pi_s$ was calculated using the 2006 MCAS. $\pi_s$ represents the exposure of Mexican sending regions to LAWA, so the negative relationship implies decreased emigration from Mexican sources whose migrant networks were more exposed to Arizona. Panel (a) shows the relationship at the Mexican state level, and Panel (b) shows the relationship at the municipio level. The regression lines are based on the regression model in (2), with no controls. See Table 4 for alternative specifications.
Table 1: Stock of Valid MCAS, Mexican-born Population, and Unauthorized Mexican-born Population

<table>
<thead>
<tr>
<th>Year</th>
<th>Valid MCAS</th>
<th>Mexican-born Population</th>
<th>Unauthorized Mexican-born Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>4,623,531</td>
<td>11,940,336</td>
<td>6,150,000</td>
</tr>
<tr>
<td>2012</td>
<td>4,506,012</td>
<td>11,813,907</td>
<td>5,850,000</td>
</tr>
<tr>
<td>2013</td>
<td>4,521,100</td>
<td>11,762,136</td>
<td>5,600,000</td>
</tr>
</tbody>
</table>

The “Valid MCAS” column is based on MCAS tabulations provided by Institute for Mexicans Abroad (Instituto de Los Mexicanos en el Exterior, IME). Each row reports the number of cards issued in the preceding 5 years, which are still valid as of the year listed. The “Mexican-born Population” column is estimated using American Community Survey data. We average population estimates for the year listed and the prior year to center the estimates around the beginning of the listed year, corresponding to the timing of MCAS card validity. The “Unauthorized Mexican-born Population” column reports figures from Pew Research Center Reports (Passell and Cohn 2014, Gonzalez-Barrera 2015).
Table 2: Characteristics of Matrícula Applicants vs. Overall Mexican-Born Population

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Matrícula Applicants</th>
<th>Overall Mexican-born Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.57</td>
<td>0.55</td>
</tr>
<tr>
<td>Female</td>
<td>0.40</td>
<td>0.45</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>0.48</td>
<td>0.30</td>
</tr>
<tr>
<td>30-39</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>40-49</td>
<td>0.13</td>
<td>0.20</td>
</tr>
<tr>
<td>50-54</td>
<td>0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>55+</td>
<td>0.05</td>
<td>0.12</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary or less</td>
<td>0.34</td>
<td>0.23</td>
</tr>
<tr>
<td>Lower secondary / vocational</td>
<td>0.36</td>
<td>0.23</td>
</tr>
<tr>
<td>High school</td>
<td>0.23</td>
<td>0.32</td>
</tr>
<tr>
<td>College+</td>
<td>0.07</td>
<td>0.21</td>
</tr>
<tr>
<td>Arrival Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before 1990</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>1990-1994</td>
<td>0.16</td>
<td>0.15</td>
</tr>
<tr>
<td>1995-1999</td>
<td>0.28</td>
<td>0.20</td>
</tr>
<tr>
<td>2000-2004</td>
<td>0.38</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Reports the individuals with each of the characteristics listed in the first column. The “Matrícula Applicants” column is based on the Survey of Mexican Migrants (Suro 2005). The “Overall Mexican-born Population” column is based on the 2005 American Community Survey (ACS).
Table 3: Estimated Effect of LAWA on Return Migration Rates

<table>
<thead>
<tr>
<th></th>
<th>Change in log return migration rate</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>unweighted</td>
<td>weighted by 2000</td>
<td>control for outliers</td>
<td>unweighted</td>
<td>weighted by 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>population</td>
<td>(2)</td>
<td>outliers</td>
<td>(3)</td>
</tr>
<tr>
<td>Arizona’s share of migrants in 2006</td>
<td>0.56**</td>
<td>0.54*</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.22)</td>
<td>(0.29)</td>
<td>(0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.34***</td>
<td>1.37***</td>
<td>1.33***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.022</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A. Mexican State level Estimates

Panel B. Mexican Municipio level Estimates

Relates the change in log return migration rate from 2005 to 2010 to Arizona’s initial share of migrants from the Mexican source region. Positive regression coefficients indicate that regions with a stronger initial connection to Arizona experience a larger increase in return migration following the implementation of LAWA. The return migration rate is calculated using data from the 2000 and 2010 Mexican Census and the 2005 Conteo, as the number of return migrants during 2000-2005 or 2005-2010, divided by the source region’s initial population. Arizona’s initial migrant share is calculated using the 2006 MCAS. Panel A examines Mexican state-level source regions, while Panel B examines municipio-level sources. Columns (1) and (4) estimate unweighted (equally weighted) regressions across source regions. Columns (2) and (5) weight sources by their 2000 population, and columns (3) and (6) use robust regression to reduce the influence of high leverage outliers. In Panel B, columns (4)-(6) control for Mexican state fixed effects. Heteroskedasticity robust standard errors are shown in parentheses. In Panel B, the standard errors are clustered by Mexican state. *** p<0.01, ** p<0.05, * p<0.1.
Table 4: Estimates of the Effect of LAWA on Emigration Rates

<table>
<thead>
<tr>
<th></th>
<th>Change in log of emigration rate</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unweighted</td>
<td>weighted by 2005 population</td>
<td>control for outliers</td>
<td>unweighted</td>
<td>weighted by 2005 population</td>
<td>controls for</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Panel A. Mexican State level Estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona’s share of migrants in 2006</td>
<td>-1.02**</td>
<td>-0.89*</td>
<td>-0.98</td>
<td>(0.38)</td>
<td>(0.51)</td>
<td>(1.25)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.23*</td>
<td>-0.36**</td>
<td>-0.23</td>
<td>(0.13)</td>
<td>(0.17)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Observations</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Panel B. Mexican Municipio level Estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arizona’s share of migrants in 2006</td>
<td>-0.82**</td>
<td>-0.60</td>
<td>-0.83***</td>
<td>-1.16**</td>
<td>-1.74**</td>
<td>-1.26***</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.04</td>
<td>-0.57***</td>
<td>-0.05</td>
<td>0.41***</td>
<td>0.25***</td>
<td>0.47</td>
</tr>
<tr>
<td>State FE</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>1,752</td>
<td>1,752</td>
<td>1,752</td>
<td>1,752</td>
<td>1,752</td>
<td>1,752</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.18</td>
<td>0.25</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Relates the change in log emigration rate from 2005 to 2010 to Arizona’s initial share of migrants from the Mexican source region. Negative regression coefficients indicate that regions with stronger initial connection to Arizona experience a larger decline in emigration following the implementation of LAWA. The emigration rate is calculated using data from 2010 Mexican Census, as the number reporting emigration in 2005 or 2010, divided by the source region’s population in that year. Arizona’s initial migrant share is calculated using the 2006 MCAS. Panel A examines Mexican state-level source regions, while Panel B examines municipio-level sources. Columns (1) and (4) estimate unweighted (equally weighted) regressions across source regions. Columns (2) and (5) weight sources by their 2000 population, and columns (3) and (6) use robust regression to reduce the influence of high leverage outliers. In Panel B, columns (4)-(6) control for Mexican state fixed effects. Heteroskedasticity robust standard errors are shown in parentheses. In Panel B, the standard errors are clustered by Mexican state. *** p<0.01, ** p<0.05, * p<0.1.
A Appendix

A.1 MCAS Destination Geography

Figure A-1 shows the 75 destination areas that can be separately observed in the MCAS tabulations by combining information at the U.S. state level and the consular area levels. Note that areas with larger populations tend to have finer geographic detail.

Figure A-1: Map of U.S. Destinations Identifiable in MCAS Tabulations

This map shows the smallest geographic partition of the U.S. that can be constructed using MCAS tabulations by combining tabulations at the U.S. state and consular area levels.
### A.2 MCAS Renewals and Issuances

Table A-1 reports the number of MCAS cards issued each year and the share that were new issuances vs. renewals. On average, slightly fewer than 1,000,000 of cards were issued each year from 2006 to 2013, with more than 7,000,000 millions cards issued over the entire period. The vast majority of cards issued were new issuances rather than renewals, which mitigates concerns about double-counting individuals receiving renewed cards within 5 years of the original card’s issuance.

Table A-1: MCAS Cards Issued

<table>
<thead>
<tr>
<th>Year</th>
<th>Cards Issued</th>
<th>Share New Issuances</th>
<th>Share Renewals</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>945,065</td>
<td>99.81%</td>
<td>0.19%</td>
</tr>
<tr>
<td>2007</td>
<td>906,802</td>
<td>99.78%</td>
<td>0.22%</td>
</tr>
<tr>
<td>2008</td>
<td>1,007,778</td>
<td>99.73%</td>
<td>0.27%</td>
</tr>
<tr>
<td>2009</td>
<td>921,565</td>
<td>99.63%</td>
<td>0.37%</td>
</tr>
<tr>
<td>2010</td>
<td>842,321</td>
<td>98.98%</td>
<td>1.02%</td>
</tr>
<tr>
<td>2011</td>
<td>827,546</td>
<td>98.05%</td>
<td>1.95%</td>
</tr>
<tr>
<td>2012</td>
<td>921,890</td>
<td>97.67%</td>
<td>2.33%</td>
</tr>
<tr>
<td>2013</td>
<td>951,570</td>
<td>97.04%</td>
<td>2.96%</td>
</tr>
<tr>
<td>Total</td>
<td>7,324,537</td>
<td>98.84%</td>
<td>1.16%</td>
</tr>
</tbody>
</table>

Authors’ calculations using data on MCAS cards issuances and the share of new issuances and renewals, provided by the Institute for Mexicans Abroad (Instituto de los Mexicanos en el Exterior, IME).
A.3 Comparisons between EMIF and Census data sources

This section examines the source and destination distributions for potential migrants, reported in the *Encuesta sobre Migración en la Frontera Norte* (EMIF).

Figure A-2 compares the share of Mexican migrants planning to migrate to each U.S. state, as reported in the EMIF, to the distribution of Mexican arrivals in each U.S. state, as reported in the ACS, during 2006-2013. There are substantial differences for many U.S. destination states, reflected in the dissimilarity index indicating that 32 percent of the planned migrants from the EMIF would need to be reallocated across U.S. states to exactly match the realized destination distribution in the ACS. Note that the EMIF performs far worse than the MCAS (Figure 1, dissimilarity = 0.08) and the ENADID (Figure 3 Panel (a), dissimilarity = 0.18). Because of this poor performance in matching the distribution of U.S. destinations, we do not examine the EMIF in detail in the main text.

Figure A-3 compares the share of migrants from each Mexican state planning to migrate to the U.S., as reported in the EMIF, to the distribution of emigrants across Mexican State, as reported in the Mexican Census. In this case, the EMIF has a dissimilarity index of 0.17, performing similarly to the MCAS (Figure 2 Panel (a), dissimilarity = 0.16) and the ENADID (Figure 3 Panel (a), dissimilarity = 0.22).
The figure plots the distribution of Mexican-born individuals across U.S. destinations. Each point represents the natural log of the share of individuals in each dataset selecting the relevant U.S. state as a destination. The ACS sample includes Mexican-born individuals who in the 2013 ACS report arriving in the U.S. during 2006-2013. The EMIF sample includes individuals who report intending to cross into the U.S. during the same time period. The 45-degree line, which would indicate perfect agreement between the two data sources, is shown for reference. The dissimilarity index, shown in the lower left corner, is defined in [1] and is interpreted as the share of individuals that would need to be reallocated to make the two datasets’ distributions match exactly.
The figure plots the distribution of Mexican source states for migrants to the U.S. Each point represents the natural log of the share of U.S. migrants in each dataset from each Mexican state. The Mexican Census sample includes individuals who moved to the U.S. during the five year year period from June 2005 through June 2010. The EMIF sample includes individuals from Mexican sending states who intended to cross into the U.S. during 2005-2010. The 45-degree line, which would indicate perfect agreement between the two data sources, is shown for reference. The dissimilarity index, shown in the lower left corner, is defined in (1) and is interpreted as the share of individuals that would need to be reallocated to make the two datasets’ distributions match exactly.
A.4 ENADID Sample Size Limitation

Figure [A-4] provides further evidence for the small sample size issue in the ENADID. We plot the empirical cumulative distribution function (CDF) for the number of (unweighted) individual migrant observations for each Mexican source state for the ENADID in Panel (a) and the MCAS in Panel (b). Both figures show a reference line at 51 observations. The intersection of this reference line with the empirical CDF reports the share of Mexican source states for which we observe fewer than 51 migrants, which is the minimum number needed to have no empty source-destination cells when using U.S. states as destinations. Panel (a) reveals that about 80 percent of Mexican source states in ENADID have fewer than 51 observations while Panel (b) reveals that all source states in the MCAS have more than 51 observations. In fact, the majority of source states have more than 100,000 observed migrants in the MCAS data.
Figure A-4: Comparison of the Distribution of Number of Observations for each Mexican State: ENADID vs. MCAS data

Each figure plots the empirical cumulative distribution function for the number of (unweighted) individual migrant observations for each Mexican source state. Panel (a) uses the ENADID, and Panel (b) uses the MCAS. The reference line in each figure is 51, which is the minimum number of observations needed to have no empty source-destination cells when using U.S. states as destinations.
A.5 Detailed Destination Distributions for Different Sources

Figure A-5 shows the distribution of U.S. residences for MCAS cardholders born in Michoacán, and Figure A-6 provides the distribution for all cardholders born anywhere in Mexico. These two distributions are used to calculate the differences shown in Figure 5 in the main text.

Figures A-7 and A-8 provide an example of the variation in destinations selected by migrants from two municipios within the same state. Migrants from Tiquicheo (Figure A-7) and from Ciudad Hidalgo (Figure A-8) choose substantially different destinations, even though both municipios are in Michoacán.

These examples demonstrate that for many municipios the destination distribution of their state is not be a good proxy. To reinforce this conclusion, Figure A-9 provides state-by-state differences in shares between Tichicheo and all migrants from Michoacán, revealing a substantial divergence.
Figure A-5: Distribution of MCAS Card Issuances for Migrants Born in Michoacán

This map shows the share of MCAS identity cards issued to migrants born in Michoacán who had current addresses in each US state. The data source is the universe of MCAS identity cards issued during the 2006-2010 time period.
Figure A-6: National Distribution of MCAS Card Issuances

Notes: This map shows the share of MCAS identity cards issued to migrants with addresses in each US state. The data source is the universe of MCAS identity cards issued during the 2006-2010 time period.
This map shows the destination distribution of the share of MCAS identity cards issued to migrants born in the municipio of Tiquicheo (shown in bright blue) who had current addresses in each US state. The data source is the universe of MCAS identity cards issued during the 2006-2010 time period.
Figure A-8: Distribution of MCAS Card Issuances for Migrants Born in Hidalgo

This map shows the destination distribution of the share of MCAS identity cards issued to migrants born in the municipio of Hidalgo (shown in bright blue) who had current addresses in each US state. The data source is the universe of MCAS identity cards issued during the 2006-2010 time period.
Figure A-9: Difference in Distribution of MCAS Card Issuances for Migrants Born in Tiquicheo Compared to Michoacán

This figure shows the destination distribution of the state-by-state difference between the share of individuals born in Tiquicheo (a Mexican municipio within Michoacán) shown in Figure A-7 and the share of individuals born in the Mexican state of Michoacán shown in Figure A-5. The data source is the universe of MCAS identity cards issued during the 2006-2010 time period.