MEXICAN IMMIGRATION AND SELF-SELECTION: NEW EVIDENCE FROM THE 2000 MEXICAN CENSUS[*]

Pablo Ibarraran
Inter-American Development Bank

Darren Lubotsky
University of Illinois at Urbana-Champaign

PRELIMINARY DRAFT
PLEASE DO NOT CITE

January 2005

Abstract

We use data from the 2000 Mexican Census to examine how the education and socioeconomic status of Mexican immigrants to the United States compares to that of non-migrants in Mexico. Our primary conclusion is that migrants tend to be less educated than non-migrants. This finding is consistent with the idea that the return to education is higher in Mexico than in the United States, and thus the wage gain to migrating is proportionately smaller for high-educated Mexicans than it is for lower-educated Mexicans. We also find that the degree of negative selection of migrants is stronger in Mexican counties that have a higher return to education.

[*] This draft is prepared for the NBER Mexican Immigration Conference, February 11th and 12th, 2005, Cambridge, Massachusetts. For helpful comments and discussions we thank participants at the NBER Mexican Immigration Pre-Conference, and Richard Akresh, Todd Elder, Kevin Hallock, and Roger Koenker.

I. Introduction

We use data from the 2000 Mexican Census to examine how the educational attainment of Mexican migrants to the United States compares to the educational attainment of those who remain in Mexico. We present a version of the standard economic model of migration that predicts lower-educated Mexicans are more likely to migrate to the United States than higher-educated Mexicans. Moreover, we expect there to be substantial variation in the degree of migrant selectivity throughout Mexico: Areas within Mexico that have high returns to education will tend to attract more highly educated Mexicans and provide a greater incentive for low-educated Mexicans to move to the United States. By contrast, lower-educated Mexicans will tend to remain in those areas within Mexico that have a relatively lower return to education. Migration from these areas will tend to be more balanced between higher and lower-educated Mexicans, or may even favor high educated Mexicans.

Alternative theories of migration posit that wage differences between countries may not be important determinants of the magnitude and skill-composition of migratory flows, and factors such as migration costs, community social capital, migration networks, and access to credit markets may be more important. Some of these theories predict that Mexican migrants will be positively selected; that is, they will be more skilled than non-migrants. Our primary goal is to accurately assess whether migrants are in fact positively or negatively selected as a first step in determining the relative importance of wage differences, returns to human capital, and other influences on Mexico-U.S. migration patterns.

Knowing whether Mexicans tend to come from the bottom or the top of the Mexican socioeconomic scale has important implications for a number of research and policy questions. Perhaps most importantly, migration may have profound effects on the Mexican labor force and, though remittances, on the economic well-being of families in Mexico. In one view, if migration responds to differences in the return to skills between countries and migrants are largely

composed of less-skilled Mexicans, then migration will tend to equalize economic status among Mexican families. Moreover, economic development in Mexico may lead to a reduction in migratory flows and a relative increase in the skills of Mexican migrants to the United States. On the other hand, if household wealth or access to credit markets are important pre-conditions for migration, migrants will tend to be drawn from the upper half of the Mexican skill distribution and economic development may lead to increased migration and increased inequality within Mexico.

U.S. immigration policy is routinely criticized for encouraging too many low-skilled immigrants and too few high skilled immigrants. A better understanding of the determinants of the stock of migrants to the U.S. is critical for evaluating the likely effects of alternative policies. For example, the fear that increased welfare generosity or increases in the minimum wage will encourage low-skilled migration is more realistic if low-skilled Mexicans indeed do respond to earnings differences between Mexico and the United States. On the other hand, English language programs and other policies that may increase the returns to skills may be more likely to increase migration among higher skilled Mexicans.

Finally, studies of immigrants' performance in the U.S. labor market typically compare immigrants' earnings to that of native-born workers.[1] While this comparison is certainly interesting and important, it does not tell us the extent to which the well-being of immigrants improved as a result of their migration. A better understanding of the socioeconomic status of Mexican migrants and their families back in Mexico will help us to put the immigrant labor market experience in the U.S. in a wider perspective.

Our main finding is that less-skilled Mexicans are more likely than higher skilled Mexicans to migrate to the United States. Moreover, consistent with the predictions of the theoretical model, the degree of negative selection among migrants is larger in counties within the Mexican states where migrants typically originate that have higher returns to education. We also

[1] For example, Trejo (1997) studies the earnings of Mexicans in the U.S. labor market.

find that Mexican immigrants in the 2000 U.S. Census are older and significantly better skilled than migrants in the 2000 Mexican Census. Though part of this discrepancy is likely caused by the particular sampling procedure of the Mexican Census, part is also likely caused by the undercount of young, largely illegal Mexican immigrants and over-reporting of education in the U.S. Census.

The paper proceeds as follows: in the next section we discuss the standard theoretical framework to analyze migration and selection, and we review the literature on education and self-selection of Mexican migrants. In Section III we describe the 2000 Mexican Census and compare its coverage of migrants with that in the 2000 U.S. Census. Section IV compares the level of education among migrants and non-migrants. Section V investigates the relationship between the degree of migrant selection and local returns to education. Finally, Section VI concludes.

II. Theory and existing evidence

We begin with a standard migration model in which Mexicans compare their potential earnings in Mexico with their potential earnings in the United States net of moving costs.[2] Let the log earnings of individual i who lives in Mexican county c be given by

$$\log(w_{ic}) = \alpha_c + \beta_c S_i \quad (1a)$$

where S_i is the level of schooling completed by the individual, β_c is the return to schooling in county c, and α_c captures the level of earnings in county c. If the individual were to move to the United States, his log earnings would be determined by

$$\log(w_{iu}) = \alpha_u + \beta_u S_i \quad (1b)$$

where β_u is the return to education faced by Mexican immigrants in the United States. Our formulation of the model assumes there is variation at the county level in the average level of

[2] The model is similar to those in Borjas (1987, 1991, and 1999) and Chiquiar and Hanson (2004).

earnings and the returns to schooling within Mexico, but there is a single rate of return in the United States. We assume these rates of return are exogenously given.

A person migrates to the United States if the wage gain plus any non-pecuniary gains outweigh the costs of migration. Denote by C_{ic} the migration costs net of any non-pecuniary gains for person i moving from county c to the United States. The person migrates if $\log(w_{iu} - C_{ic}) \approx \log(w_{iu}) - \pi_{ic} > \log(w_{ic})$, where $\pi_{ic} = C_{ic} / w_{ic}$ is the time-equivalent net costs of migration. The wage gain to individual i were he to move to the United States from county c is given by

$$\begin{aligned} G_{ic} &= (\alpha_u + \beta_u S_i) - (\alpha_c + \beta_{cu} S_i) \\ &= (\alpha_u - \alpha_c) + S_i(\beta_u - \beta_c) \end{aligned} \quad (2)$$

The migration decision can therefore be expressed as a comparison of the wage gain G_{ic} to the time equivalent net migration costs π_{ic}. The person migrates if $G_{ic} > \pi_{ic}$, which is equivalent to

$$(\alpha_u - \alpha_c - \pi_{ic}) + S_i(\beta_u - \beta_c) > 0 \quad (3)$$

Equation 3 highlights two predictions about the selectivity of Mexican migrants to the United States. First, the wage gain to migrating, G_{ic}, will either rise or fall with additional education, depending on whether the return to education is higher in the United States or in Mexico, which implies that migrants will be drawn from either the upper tail or the lower tail of the distribution of education. When the return to education is higher in Mexico than in the U.S., $(\beta_u - \beta_c) < 0$, and the wage gain from migrating to the United States is larger for low-educated Mexicans than it is for higher educated Mexicans.

The second predictions is that among low-educated Mexicans for whom the wage gain to migrating is positive, the wage gain is larger in Mexican counties with relatively larger returns to schooling. That is, Mexican migrants will tend to be negatively selected from all counties in Mexico that have a rate of return to schooling higher than that in the United States, but the selection will be stronger in counties with relatively higher rates of return.

Though the model captures the essential idea behind wage differences as a driving force behind migration incentives, it contains a number of simplifications. Perhaps most importantly, the rate of return to education in a Mexican county is not necessarily exogenously given, as we have assumed. Instead, it is likely to be jointly determined with the skill composition of migrants moving from the county to the United States, and with the skill composition of internal migration within Mexico. The model also ignores aspects of skills besides education and the possibility of back-and-forth migration between Mexico and the United States. Finally, some recent work stresses the importance of networks and social capital in the migration process. [3] One can view these institutions as either influencing the net costs of migration, C_{ic}, the level of earnings in the United States, α_u, or the return to education in the United States, β_u, for some migrants more than others.

The predictions about migrant selectivity are driven by wage differences between Mexico and the United States, but ignore the influence of migration costs, a recent source of controversy in the literature. Chiquiar and Hanson (2004) suggest that time-equivalent migration costs may fall with education and this may lead migrants to be drawn from the middle or upper tail of the Mexican skill distribution, even if the wage gain is largest among the least skilled Mexicans. For example, fixed costs of migrating will translate into a smaller time-equivalent cost for high-wage migrants than for low-wage migrants. There may also be higher borrowing cost among low-income Mexican families than among high-income families. On the other hand, migration costs may be higher for better skilled workers if they require legalized status to practice their profession in the United States. In any event, little is known about the magnitude or source of migration costs.

Though the literature on Mexican immigration is vast, there is very little that focuses on the selectivity of migration. Chiquiar and Hanson (2004) compare Mexicans in the 1990 and 2000

[3] For example, see Durand, Massey, and Zenteno (2001); de Janvry, Sadoulet, and Winters (2001); Massey and Singer (1998); and Munshi (2003).

U.S. Censuses to non-migrant Mexicans in the 1990 and 2000 Mexican Censuses. They conclude that migrants, if they were to return to Mexico, would tend to fall in the middle or upper part of the Mexican wage distribution, which suggests that factors other than wage differences play an important role in shaping Mexican migration. In a similar type of analysis, Cuecuecha (2003) compares Mexicans in the 1994 United States Current Population Survey with Mexicans in the 1994 Encuesta Nacional de Ingreso y Gasto de los Hogares, an income and consumption survey, and also concludes that positive selection takes place within Mexico.

Prior to the release of the 2000 Mexican Census, which we will describe in more detail below, a primary source of data on both Mexican residents and migrants to the United States is the Mexican Migration Project. Orrenius and Zavodny (2004) use this data to examine how various factors influence the selectivity of migrants over time. Among their findings are that improvements in U.S and Mexican economic conditions lead to increased negative selection of migrants, but stricter border enforcement, coupled with deteriorating conditions within Mexico, lead to increased positive selection. Their descriptive statistics suggest that overall migrants come from the middle of the distribution of education.

In the remainder of the paper we use data from the 2000 Mexican Census and the 2000 U.S. Census to compare the educational attainment of migrants and non-migrants. In doing so, we also attempt to shed light on how coverage of Mexican immigrants differs across the two data sources.

III. Description of the Mexican Census data and its coverage of Mexican migrants

With the right data, comparing the skills of migrants to non-migrants in Mexico is straightforward: The ideal dataset would contain information on all Mexicans at a point in time, indicators for which Mexicans moved to the United States during some subsequent time period, and a set of exogenous measures of each individual's skill and the return to skill in their local

area. Unfortunately, this ideal dataset does not exist. Instead, past researchers have relied on the alternative data sources described above.

We take a new approach and use the 2000 Mexican Census to compare the characteristics of Mexican migrants and non-migrants, and their families. This data source has some important limitations compared to the ideal dataset described above, but it also overcomes some limitations of other datasets used in prior work. The Census was conducted in February 2000. Household heads were asked to list all current members of the household and to also list any current or past household member who had lived abroad during the preceding five years. A relatively large amount of economic and demographic information is collected about current household members. A much more limited amount of information is collected on the migrants, including their age, gender, Mexican state of origin, month and year of most recent departure, destination country, and current country of residence. About 16 percent of migrants had returned to Mexico and the Census records the month and year of their return.[4] The data consist of a 10% sample of the Mexican population. Household weights are provided to adjust for non-response. There are 2,312,035 households in the sample, consisting of 10,099,182 total persons who live in Mexico.

There are three primary limitations of the Mexican Census relative to our ideal dataset:

1. We do not have key socioeconomic information about the migrants themselves. In particular, we do not know their educational attainment or labor market success in Mexico prior to moving to the United States. We also don't know migrants' relationship to the household members in Mexico.
2. We do not have clearly exogenous indicators of socioeconomic status. Non-migrant Mexicans may reduce their labor supply if they receive large remittances from family

[4] Thus, a household member could be listed as both a current household member and as an international migrant if he or she had moved abroad during the past five years and had returned to the same household in Mexico. Unfortunately, the data does not directly link return migrants with current household members, or even identify if return migrants currently live in the household. At best, one could match return migrants with current household members by age and gender.

members in the United States, or may use remittances to finance additional schooling for younger household members.

3. We do not have any information about households in which all members moved to the United States. We estimate that such households account for at most 22 percent of male Mexican migrants and 48 percent of female Mexican migrants. We return to this sampling issue below.

The major advantages of these data compared to the sample of Mexican migrants in the U.S. Census are, first, that we can link migrants to their original place of residence in Mexico. This allows us to examine the influence of the local return to education on the decision to migrate among Mexicans from different points in the skill distribution. Second, there is widespread concern that the U.S. Census undercounts Mexican immigrants, and the undercount is likely to be most severe among illegal migrants and the least skilled migrants (Bean *et al* 1998).

To shed light on the relative coverage of recent Mexican immigrants enumerated in the 2000 U.S. and Mexican Censuses, we compare the population counts and characteristics of Mexicans in both data sources. It is likely that neither source will provide a fully representative sample of all Mexicans who migrated to the United States between 1995 and 2000. To recap, the two primary sources of discrepancy are, first, that data from the Mexican Census do not contain any information on whole households that moved to the U.S., while these families may be counted in the U.S. Census. Second, it is well-known that the U.S. Census undercounts immigrants, and in particular illegal immigrants. Thus, the U.S. and Mexican Censuses contain overlapping subsamples of all Mexican migrants.

In addition, there are two less-important sources of discrepancy in coverage between the two Censuses: (1) it is possible for a migrant to be counted more than once in the Mexican Census. The survey asks heads if any household member over the past five years had moved to the U.S. A migrant who was a member of multiple households during the past five years (e.g. his parents' and wife's households) could be in the Mexican data multiple times. (2) The U.S. Census

was taken on April 1st, two months after the Mexican Census. Migrants who moved in February or March of 2000 may be in the U.S. Census but not show up as migrants in the Mexican Census. If the migration flows during these two months is equal to average flow between 1995 and 2000, this discrepancy will lead to an increase in the U.S. Census count of about 1/30th, or 3 1/3 percent, over five years, relative to the Mexican Census count.[5]

With these differences in mind, in Table 1 and Figures 1a and 1b we compare Mexican migrants in the Mexican and U.S. Census. Panel A of Table 1 shows estimates of the migrant population taken from the Mexican Census. Excluding the 15 percent of Mexican migrants who are reported to have returned to Mexico by February 2000, there are 115,760 male migrants aged 16 or older and 33,516 females in the data. Using the household weights provided in the Mexican Census, these observations correspond to a population estimate of 925,587 males and 296,011 females. The average migrant is 26 years old.

In panel B we show analogous estimates of the Mexican immigrant population in the United States from the 2000 U.S. Census. This sample includes all people who report that they came to the U.S. between 1995 and April 2000, or reported that they lived in Mexico in April 1995. There are 62,408 males and 41,403 females in the data, corresponding to population estimates of 1,324,762 and 880,594. The average male is 28 years old and the average female is 30 years old. Thus, the male and female populations in the Mexican Census are about 70 and 34 percent of the size of the populations in the U.S. Census.

The right-side of Panel B presents population estimates from the U.S. Census that exclude migrants who report themselves as married with spouse present. Since Mexican married couples in the United States seem most likely to have migrated as a whole household, they are most likely to be missing from the migrant population in the Mexican Census. The population estimates for the remaining migrants in the U.S. Census are 1,033,060 men and 459,051 women.

[5] Many Mexican immigrants return to Mexico during the winter, before returning to the U.S. We do not know the size of this flow, but if it is large, this difference between sampling dates in Mexico and the U.S. could lead to a larger discrepancy that 3 1/3 percent.

The total population estimates from the Mexican Census correspond to 90 and 64 percent of these population estimates.

A different way to compare the coverage rates in the two data sources is to plot the population estimates of disaggregated subgroups. Figure 1a is a plot of the population estimate from the Mexican Census against the estimate from the U.S. Census for men in two-year age groups from 16 to 50 and five-year age groups from 50 to 70. Figure 1b is the analogous plot for women.[6] The dashed 45 degree line represents an equal population estimate in the two data sources. The solid line is a fitted regression line.

Figure 1a shows different patterns for three distinct groups. First, that there are more migrants aged 16 to 19 in the Mexican Census than enumerated in the U.S. Census. Members of this group are more likely than older migrants to be in the U.S. illegally, less likely to have set up permanent roots in the United States, and hence more likely to be undercounted in the U.S. Census. Men in this group are also less likely to be married and hence still have household members in Mexico who answered the Mexican Census. Second, men aged 20 to 31 are overrepresented in the U.S. Census compared to the Mexican Census. Third, men aged 32 to 75 are also overrepresented in the U.S. Census, though the degree of overrepresentation is less than that for men aged 20 to 31.

Figure 1b shows a similar pattern of coverage across the three age groups of women, except that the U.S. Census contains many more female migrants than does the Mexican Census. This is certainly because many women only migrate as part of a whole household, and thus are not enumerated at all in the Mexican Census.

To gauge the magnitude of differences in the age distribution of Mexican immigrants in the U.S. and Mexican Censuses, Table 2 shows the how the population of recent migrants differs in the two Censuses and a third, smaller dataset, the migration module from the 2002 National

[6] The population estimates of children under age 16 are considerably smaller in the Mexican Census than in the U.S. Census, almost certainly because most children move only when the whole household moves. We thus exclude children from our population comparisons.

Employment Survey (ENE). As we saw in Table 1, the sample from the U.S. Census is about a year older than the sample from the Mexican Census.[7] The right two columns of the table show how high school graduation rates of Mexican immigrants in the U.S. Census and the fraction of immigrants in the U.S. illegally vary by age. The high school graduation rate from the U.S. Census is 27.1% and the fraction in the U.S. illegally is 79.4%. If we reweight the Mexican immigrants in the U.S. Census to have the same distribution across the five age categories as migrants in the Mexican Census, the high school graduation rate falls by 1.1 percentage points to 26.0%. Like the Mexican Census, the National Employment Survey asks household members in Mexico whether any other members have recently moved the United States. Migrants' legal status is reported by the household members in Mexico. Though this sample is older than the sample from either Census, it does indicate that older immigrants are substantially less likely to be in the United States illegally.

From these comparisons we conclude that migrants in the Mexican Census make up a fairly representative sample of the large group of men who migrate to the United States, and for this reason we focus most of the remainder of our analysis on men's migration decisions. This data is less well-equipped to provide data on entire households that move to the United States, a group that may be more educated than the typical Mexican migrant. The Mexican Census may be a better source of data than the U.S. Census on migrants aged 16 to 19, a group that is disproportionately less educated than the typical Mexican migrant.

IV. Differences in educational attainment between migrants and non-migrants

A direct comparison of the educational attainment of migrants and non-migrants in the Mexican Census is not possible because education was not recorded for the migrants. We instead pursue several alternative strategies: first, we compare educational attainment of non-migrants in

[7] These numbers differ slightly from those in Table 1 because we have restricted the sample to men aged 16 to 54.

the Mexican Census to migrants in the 2000 U.S. Census. Echoing the findings of Hanson and Chiquiar (2004), we also find that Mexicans in the U.S. Census appear to be more educated than non-migrants in Mexico. We next turn to two comparisons of educational attainment using only the Mexican Census: First, we compare the educational attainment of non-migrant Mexicans who live in household that had a migrant to the education of non-migrants that live in households without any migrants. Second, we use other information available in the Mexican Census to develop a predicted level of education for both migrants and non-migrants in Mexico.

Most Mexicans have 6, 9, 12, 16, or 17 years of education, corresponding to finishing primary school, secondary school, high school, and college. Column 1 of Table 3 shows that 45.3 percent of Mexican men aged 16 and over have six or fewer years of schooling and 76.9 percent have not completed 12 years of schooling. Column 4 of Table 3 shows the distribution of educational attainment among recent Mexican immigrant men in the 2000 U.S. Census. 73.3 percent of these Mexican immigrants have not completed high school. Among the remaining quarter or so of each population, recent Mexican immigrants in the U.S. are more likely to have a high school degree, but less likely to have a college degree, than Mexicans in Mexico. These conclusions hold when the respective samples are restricted to men aged 16 to 35.

Two concerns with the comparison of educational attainment between the Mexican and U.S. Censuses are the undercount of Mexican migrants in the U.S. Census and the possible over-reporting of education among Mexicans in the U.S. Census. In the last section we detailed differences in the age distribution of migrants in the two Censuses that indicate the U.S. Census undercounts younger migrants. A related worry is that it undercounts low-educated migrants of all ages. Both forms of undercounting migrants will lead to an overstatement of the educational attainment of Mexican migrants in the U.S. Census, which may make it appear as though migrants in the U.S. Census are better educated than non-migrants in the Mexican Census.

Migrants in the U.S. Census may also appear better educated than non-migrants if they tend to over-report their education. We do not have a method to directly test for a reporting bias

among Mexican immigrants in the U.S. Census, but a suggestive piece of evidence that Mexican immigrants in the U.S. may overstate their educational attainment (or understate their age) is that 7.2 percent of 16 and 17 year old Mexicans claim to have a high school degree or more education, compared to 3.6 percent of American-born men. In both countries a person would typically be in their third and final years of high school at ages 16 and 17.[8]

We next turn to a comparison of educational attainment of non-migrant Mexicans and migrants' family members who remain in Mexico. The advantage of this type of comparison is that we are able to use a common data source for both groups and we have self-reported education for both groups. We begin our comparison in Table 4 with a simple tabulation of the maximum completed years of education over all household members. Migrant families are more likely than non-migrant families to have zero to 11 years of education, while non-migrant families are more likely to have 12 or more years of education. Non-migrant families have, on average, about 0.8 years more schooling than migrant families.

There are two immediate problems with concluding from these data that migrants tend to come from lower educated households. First, migrants tend to be men aged 16 to 35, a group that tends to have high educational attainment within Mexico. Thus, migrant households are likely to be missing their "high education" member, while non-migrant households contain him. This would lead us to understate the education of migrant households. Second, if children tend to be the highest educated member of migrant households, while adults tend to be the highest educated member of non-migrant households, then the maximal education in the household may be a poor barometer of the overall economic well-being of the household.

One simple way to address these concerns is to compare the highest educated woman across households. Since about 75 percent of migrants are men, measurement of household educational attainment of women in Mexico is much less affected by the absence of migrants.

[8] Our tabulations from the U.S. Census include allocated educational attainment. Excluding allocated data lowers the average educational attainment of Mexican migrants, though they still appear to be slightly better educated than non-migrants in the Mexican Census.

The middle panel in Table 4 shows that migrants also tend to come from households with less educated women. The highest educated woman in 55 percent of migrant families has eight or fewer years of education, while 47 percent of non-migrant families fall in that range. The right-hand panel of Table 4 restricts the sample to women aged 16 to 35, which eliminates children from the sample. These results also indicate that migrants tend to come from less educated households.

Our final and preferred method to examine the relative educational attainment of migrants and non-migrants is to generate a predicted level of education for each sample member based on their household characteristics and location. We then compare the predicted education of migrants to non-migrants.

To predict education we regress the number of years of schooling of individual i, gender j, in county c (S_{ijc}) on indicator variables for age (d_{ajc}), six indicator variables for individuals' town size, indicator variables for the number of children in the household aged zero to eight (Kid_{1jc}), indicators for the number of children 9 to 16 years old (Kid_{2jc}), indicators for the number of men aged 17 to 35 (Man_{1jc}), indicators for the number of men aged 36 and older (Man_{2jc}), indicators for the number of women aged 17 to 35 ($Woman_{1jc}$), and indicators for the number of women aged 36 and older ($Woman_{2jc}$). Formally, we run the following regression separately by gender and county, using all people aged twelve and over who are not migrants and who do not live in a migrant household:

$$S_{ijc} = \delta_1 + d_{ajc} + d_{tjc} + \delta_2 Kid_{1jc} + \delta_3 Kid_{2jc} + \delta_4 Man_{1jc} + \delta_5 Man_{2jc} + \delta_6 Woman_{1jc} + \delta_7 Woman_{2jc} + \varepsilon_{ijc} \quad (4)$$

where the δs are the regression coefficients and ε_{ijc} is the error term. The age variables include single year indicators for ages twelve to thirty, indicators for three-year groups from thirty-one to seventy, an indicator for people in their seventies, and indicator for people over eighty. The town-size indicators correspond to towns with less than 2500 people; 2500 to 14,999; 15,000 to 19,999;

20,000 to 49,999; 50,000 to 99,999; 100,000 to 499,999 and a half-million or more people. The indicator variables for the number of children, adult men, and adult women include indicators that the household contains one, two, three, or more than three of each type of person. Again, these regressions are estimated separately by gender and by county, which allows the intercept and the relationship between education and the independent variables to vary by gender and county.

Next, we use the coefficient estimates to compute the predicted education for all sample members, including non-migrants who live in migrant households and migrants themselves. This predicted education can be interpreted as an index of educational attainment or socio-economic status more generally. The regression in equation 4 and the prediction are based only on individuals' county, household-level characteristics, plus age and gender, because these are the only variables available for the migrants with which to generate a predicted education.

Comparing predicted education of migrants and non-migrants provides useful information about the relative skill level of migrants under the assumption that unmeasured determinants of schooling are equally distributed between migrants and non-migrants. We cannot directly test this assumption, but we provide suggestive evidence below that it may be reasonable.

Table 5 compares the distribution of actual and predicted education. Column 1 shows the distribution of actual years of education among men aged 16 to 54 who are not migrants and do not live in a migrant household. Column 3 shows the distribution of predicted education for this group. Not surprisingly, there is less variation in the distribution of predicted education than in the distribution of actual education (the means are equal by construction). Columns 2 and 4 show the distribution of actual and predicted education for non-migrants who live in households that also have migrants. These individuals were not included in estimation of the education regression. It is interesting to note that the average predicted education is about 0.1 years less than the average actual education, which suggests that the unmeasured determinants of schooling exert a positive, though very small, effect on this group.

Column 5 displays the distribution of migrants' predicted education. On average, the predicted education of migrants is 0.63 years less than the predicted education of non-migrants who live in non-migrant households. If the small magnitude of the discrepancy between actual and predicted education among non-migrants who live with migrants is any guide, the actual education of migrants is likely to also be less than the actual education of non-migrants.[9]

The three panels of Table 6 show differences in predicted education by region of Mexico, by age, and by town size. Two-thirds of Mexican migrant men aged 16 to 54 originate in one of the fourteen states in Central Mexico, and the migration rate in this region is 7.4%. Interestingly, the nationwide gap in predicted education is entirely driven by the difference in this region. Predicted education is approximately equal among migrants and non-migrants in the other three regions, Mexico City and the state of Mexico, the seven northern boarder states, and eight southern states. The last line in the table shows that overall non-migrants have 0.56 years more education than migrants. Tabulations in Panel B show that migrants are less educated than non-migrants in all four age groups, though the gap is larger among older Mexicans.

Finally, Panel C shows differences by town size. 42.5% of migrants come from towns with populations less than 2500, but the remaining 57.5% of migrants are fairly equally distributed among towns with populations of 2500 and larger. The migration rate is 7.8% among all towns with a population of less than 2500 and the rate tends to fall as town size increases. At the same time, average education of migrants and non-migrants tend to rise with town size. Interestingly, among people who live in towns of the same size, migrants are more educated than non-migrants. This may reflect a process in which better-educated individuals tend to migrate from smaller towns in Mexico to both larger cities and to the United States, and only the least educated people remain in small towns.

[9] Note that we cannot construct a predicted education for Mexican migrants in the U.S. Census because we do not have their Mexican household or geographic information.

Finally, in unreported tabulations we find that the median male migrant has 7.90 years of predicted education, which would place him at the 36th percentile of the distribution of predicted education of non-migrants. From the evidence presented thus far we conclude that the sample of migrants in the Mexican Census is younger and considerably less educated than the sample of migrants in the 2000 U.S. Census. Consistent with the predictions of the theoretical model, Mexican migrants in the Mexican Census are less educated than non-migrants.

V. The returns to schooling and migrant self-selection

In this section we test the prediction that the degree of selection will be larger in regions within Mexico that have relatively higher returns to schooling. Recall that in our earlier model the wage gain from migrating to the United States for a person with schooling level S_i who lives in Mexican county c is given by

$$G_{ic} = (\alpha_u - \alpha_c) + S_i(\beta_u - \beta_c) \tag{5}$$

where β_u is the return to schooling in the United States and β_c is the return to schooling in Mexican county c. A person migrates if the wage gain plus any non-pecuniary gains outweigh the costs of migration. Formally, define the indicator variable M_{ic} to equal one if person i migrates to the U.S. and zero otherwise. Then $M_{ic} = 1$ if $G_{ic} > \pi_{ic}$, where π_{ic} are time-equivalent migration costs net of any non-pecuniary gains. Alternatively, $M_{ic} = 1$ if

$$(\alpha_u - \alpha_c - \pi_{ic}) + S_i(\beta_u - \beta_c) > 0 \tag{6}$$

Lacking data on migration costs, we approximate the term $(\alpha_u - \alpha_c - \pi_{ic})$ as a function of indicators for individuals' age and either county or state of residence, and model the migration probability as

$$\Pr(M_{ic} = 1) = Age_i + d_c + S_i\lambda_1 + \hat{\beta}_c\lambda_2 + \hat{\beta}_c S_1\lambda_3 + \nu_{ic} \tag{7}$$

where Age_i is a full set of indicators for each age from 16 to 54 and d_c is either a set of state or county indicators.[10] As before, S_i is years of completed schooling. $\hat{\beta}_c$ is an estimate of the returns to schooling in county c, described below.

Our main parameter of interest is λ_3, the coefficient on the interaction between an individual migrant's education and the return to education in his county. If $\beta_u - \beta_c < 0$, then according to equation 5, schooling should have a negative influence on the wage gain to migrating, and this effect should be more negative in areas with higher returns to schooling. That is, we expect λ_3 to be negative.

We estimate the county-level returns to schooling, $\hat{\beta}_c$, by estimating a regression of the log monthly wage on years of completed schooling and a quartic in age among men aged 18 to 54. We run this model separately by county and weight each observation by the Census population weight. One clear problem is that our estimated return to education at the county level may be influenced by the relative skill levels of past migrants. If lower educated Mexicans tend to leave a county, the return to education in the county should fall. If this is an important feature of the data, it would tend to bias our regression results towards finding a positive effect of the interaction between migrants' predicted education and their local return to schooling. Lacking any credible instruments for the local return to schooling, we proceed with our OLS models.

Our estimates of the return to schooling may also be influenced by the lack of earnings data for workers in the informal sector. It is common in Mexico for workers not to receive a monetary wage, especially workers in rural areas and those working in a family business. Typically these workers have low levels of schooling and low earnings. The exclusion of these workers from our sample will likely lead us to understate the return to education in general, but may also impact the relative returns to education across areas.

[10] The main effect of county rate of return to schooling, λ_2, is not identified when county fixed effects are included in the model, but the interaction effect λ_3 is identified. Both the main and interaction effects are identified when the county fixed effects are replaced with state fixed effect.

We estimate equation 7 separately by region using a linear probability model.[11] The results are shown in Table 7 and are generally consistent with our theoretical predictions. We estimate the model without any geographic fixed effects (model 1), with state fixed effects (model 2), and with county fixed effects (model 3). In Central and Southern Mexico, the origin of nearly 80 percent of migrants, the interaction between individuals' predicted education and their county return to education has a negative and statistically significant effect on the probability of migrating to the United States. The predictions are not supported by the results from Northern Mexico or from Mexico City and the state of Mexico. The parameter estimate for the interaction effect is positive and, in the latter group, statistically significant.

To gauge the magnitude of our regression results, in Table 8 we show predicted migration probabilities derived from our regressions. These rates refer to 25 year old Mexicans, and show how migration differs between those with six and ten years of predicted education, living in Mexican counties with returns to education of either 0.06 or 0.10. In Central Mexico these levels of predicted education correspond to approximately the 30^{th} and 85^{th} percentiles and the returns to education correspond to approximately the 30^{th} and 90^{th} percentiles. At the top of this table we show the migration rate in each region, the fraction of migrants that originate in each region, and the average rate of return to education in each region. Counties in Central Mexico with a rate of return to education of 0.6 would tend to have a migration rate of 11.9% among Mexicans with six years of education, based on our results from model 1. The migration rate falls by 3.2 percentage points to 8.7% among those with 10 years of education who live in the same area. By contrast, the difference in migration rates is 5.0 percentage points in areas with a return to education of 0.10. In this scenario, migration rates are 11.3% and 6.3% for the low and high educated Mexicans.

[11] Linear probability models were considerably quicker to estimate, particularly when we included county fixed effects. We find essentially similar results using probit models.

In Central Mexico, the gap in migration rates between low and high educated Mexicans is significant even in areas with low returns to education, but the gap considerably larger in areas that have relatively high returns to education. We view this evidence as indicating strong support for the idea that local variation in the wage gap between the United States and counties throughout this part of Mexico generates economically significant variation in the incentives for different types of Mexicans to migrate to the United States.

In southern Mexico, which accounts for 15.1 percent of migration, migrants in areas with returns to education of 0.06 are slightly positively selected. The predicted migration rate is 4.8% among those with six years of predicted education and 5.4% among those with 10 years of predicted education. However, migration becomes negatively selected as the rate of return to education rises. In areas with a return of 0.10, we predict a migration of 3.3% among low-educated Mexicans and 2.8% among higher-educated Mexicans. Echoing our regression results in Table 7, higher returns to education in the Northern boarder states or in Mexico City and state do not generate an increase in the magnitude of negative selection.

VI. Conclusions

We use the 2000 Mexican Census to examine the educational attainment of Mexican migrants to the United States and their families. Our primary conclusions are that migrants tend to be less educated than non-migrants. This is consistent with the idea that the greater return to skills in Mexico provides an incentive for better-skilled Mexicans to remain in Mexico and for lower-skilled Mexicans to migrate to the United States. We also find that the degree of negative selection is magnified in Mexican counties that have relatively higher returns to skills. Finally, Mexican migrants in the 2000 United States Census appear to be better educated than non-migrant Mexicans, which may be due in part to an undercount of younger, illegal, and low-skilled Mexicans.

VII. References

Bean, Frank D., Rodolfo Corona, Rodolfo, Tuirán, and Karen A. Woodrow-Lafield, 1998, "The Quantification of Migration between Mexico and the United States," in *Migration Between Mexico and the United States: Bi-National Study*, Mexican Ministry of Foreign Affairs & U.S. Commission on Immigration Reform.

Borjas, George, 1987, "Self-Selection and the Earnings of Immigrants," *American Economic Review*, Vol. 77, No. 4 (September), pp. 531-553.

Borjas, George, 1991, "Immigration and Self-Selection," in *Immigration, Trade, and the Labor Market*, John M. Abowd and Richard B. Freeman (eds.), Chicago: University of Chicago Press.

Borjas, George, 1999, "The Economic Analysis of Immigration," in Orley C. Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Volume 3A, Amsterdam: North-Holland.

Chiquiar, Daniel and Gordon Hanson, 2004, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," Working paper (August).

Corona, Rodolfo and Rodolfo Tuirán, 2001, "Fuentes Mexicanas para el Estudio de la Migración México-Estados Unidos" in Rodolfo Tuirán, *Migración México-EEUU Continuidad y Cambio*, México: CONAPO

Cuecuecha, Alfredo, 2003, "The Educational Characteristics of the Immigrants from Mexico to the U.S.," Working paper.

de Janvry, Alain, Elisabeth Sadoulet and Paul Winters, 2001, "Family and Community Networks in Mexico-U.S. Migration." *Journal of Human Resources*, Volume 36, number 1 (Winter), pp. 159-184

Durand, Jorge, Douglas Massey, and Rene Zenteno, 2001, "Mexican Immigration to the United States: Continuities and Changes", *Latin American Research Review*, Volume 36, Number 1, pp. 107-127.

Massey, Douglas and Audrey Singer, 1998, "The Social Process of Undocumented Border Crossing Among Mexican Migrants", *International Migration Review*, Volume 32, pp. 561-592.

Munshi, Kaivan, 2003, "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, Volume 118, number 2, pp. 549-597.

Orrenius, Pia and Madeline Zavodny, 2004, "Self Selection Among Undocumented Immigrants from Mexico", *Journal of Development Economics*, Forthcoming.

Trejo, Stephen, 1997, "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy*, vol. 105, No. 6, pp. 1235-1268.

Table 1: Estimates of the Mexican Immigrant Population in the U.S.

Panel A: Migrant Population Estimates from 2000 Mexican Census

	All migrants age 16 and older			All migrants age 16 and older, excluding migrants that returned to Mexico		
	All	Male	Female	All	Male	Female
Number of observations	176,448	137,910	38,538	149,276	115,760	33,516
Population estimate	1,454,690	1,111,895	342,795	1,221,598	925,587	296,011
	(3328)	(3220)	(2174)	(3054)	(2018)	(2944)
Percent female	23.6			24.2		
	(0.1)			(0.2)		
Age	26.7	26.6	27.0	26.0	26.0	26.2
	(0.03)	(0.04)	(0.08)	(0.04)	(0.04)	(0.08)

Panel B: Migrant Population Estimates from 2000 U.S. Census

	All migrants age 16 and older			All migrants age 16 and older, excluding those married with spouse present		
	All	Male	Female	All	Male	Female
Number of observations	103,812	62,409	41,403	70,752	49,048	21,704
Population estimate	2,205,356	1,324,762	880,594	1,492,111	1,033,060	459,051
	(3776)	(4500)	(4060)	(3149)	(3722)	(3077)
Percent female	40.0			30.8		
	(0.2)			(0.2)		
Age	28.7	27.8	30.0	27.2	26.4	29.1
	(0.04)	(0.05)	(0.07)	(0.05)	(0.05)	(0.10)

Notes: Population estimates are computed as the sum of the population weights in the respective surveys. Standard errors of estimates in parentheses. U.S. Census sample includes people who report that came to the U.S. between 1995 and April 2000, or reported that they lived in Mexico in April 1995.

Table 2: The Impact of Differences in Sampling on the Measured Education of Mexican Immigrants

	Age of migrants from:			HS graduation rate	Fraction Illegal
Age group	U.S. Census	Mexican Census	ENE	U.S. Census	ENE
16 to 19	17.5%	26.0%	13.9%	16.7%	86.3%
20 to 24	32.3	28.4	28.6	27.8	86.7
25 to 29	21.4	17.7	19.2	31.1	84.0
30 to 39	19.6	18.8	22.8	31.5	70.7
40 to 54	9.2	9.0	15.5	25.2	66.8
Averages	26.7	26.0	28.8	27.1%	79.4%
Sample size	60,365	134,743	2,509	60,365	2,509
HS graduation rate using age distribution from Mexican Census				26.0%	

Note: The first three columns give the fraction of recent immigrants in five age groups, computed in the 2000 U.S. Census, the 2000 Mexican Census, and the 2002 National Employment Survey (ENE). The fourth column gives the high school graduation rate tabulated from the U.S. Census and the fifth column gives the fraction in the United States illegally tabulated from the ENE data. All estimates use population weights. Sample sizes are unweighted.

Table 3: Educational Attainment of Mexican-born Men in Mexico and the United States

Years of completed education	Non-migrant Mexicans in the 2000 Mexican Census			Mexican Migrants in the 2000 U.S. Census		
	Age group			Age group		
	16 and older	16-17	16-35	16 and older	16-17	16-35
Percent of Mexicans with:						
0 to 6 years	45.3%	26.2%	31.1%	34.5%	21.7%	31.1%
7 to 11 years	31.6	71.0	43.9	38.8	71.1	41.9
0 to 11 years	76.9	97.2	75.0	73.3	92.8	73.0
12 years	9.6	2.8	12.2	17.1	6.5	18.9
More than 12 years	13.5	0.0	12.8	9.6	0.7	8.7
Sample size:	2,878,428	104,323	1,524,549	62,409	1,774	50,190

Note: All estimates use appropriate population weights. Sample sizes are unweighted. Mexican migrants in the U.S. Census are defined as those who reported that they migrated between 1995 and 2000 or who reported living in Mexico in April 1995.

Table 4: Comparison of educational attainment between migrant and non-migrant households

	Highest educated non-migrant in household		Highest educated female non-migrant in household		Highest educated female non-migrant aged 16 to 35	
Actual years of education	Migrant households	Non-migrant households	Migrant households	Non-migrant households	Migrant households	Non-migrant households
5 and under	13.1%	12.7%	23.1%	21.3%	14.3%	15.1%
6 to 8	29.3	21.1	31.5	25.5	29.4	23.7
9 to 11	32.1	29.7	26.8	26.7	34.7	32.9
12 to 15	16.5	21.5	13.3	18.0	16.4	20.4
16 or more	9.1	15.1	5.3	8.7	5.2	8.0
Average	8.9	9.7	7.6	8.3	8.4	8.9
Number of households	137,667	2,148,425	133,025	2,014,849	87,360	1,356,436

Note: A migrant household refers to a household that reports at least one migrant of either gender.

Table 5: Comparison of Actual and Predicted Education

	(1)	(2)		(3)	(4)	(5)
Actual years of education	Men in non-migrant households	Non-migrant men in migrant households	Predicted years of education	Men in non-migrant households	Non-migrant men in migrant households	Migrant men
5 and under	19.7%	26.4%	5.5 and under	12.2%	23.0%	15.6%
6 to 8	24.6	28.4	5.5 to 8.5	30.3	39.3	43.8
9 to 11	29.8	27.8	8.5 to 11.5	51.4	34.2	35.4
12 to 15	16.1	11.6	11.5 to 15.5	6.1	3.5	5.2
16 or more	9.8	5.8	15.5 and up	0.0	0.0	0.0
Mean	8.5	7.5		8.5	7.4	7.9
Sample size	2,276,862	129,733		2,276,862	129,733	134,786

Notes: Data include men aged 16 to 54 in the 2000 Mexican Census.

Table 6: Differences in Predicted Education Between Migrants and Non-migrants by Region, Age, and Town Size

Panel A: Differences by region of origin

Region	Sample size of men age 16 to 54	Migration rate	Fraction of all migrants	Average predicted education of non-migrants	Average predicted education of migrants	Difference
Central Mexico	982,906	7.4%	63.5%	7.98	7.46	0.52
Southern states	702,075	3.0%	15.1%	7.35	7.50	-0.14
Northern boarder states	405,462	2.1%	9.3%	9.05	9.02	0.03
Mexico City and state	511,081	2.2%	12.1%	9.60	9.59	0.01
All states	2,601,567	4.2%	100.0%	8.43	7.87	0.56

Notes: Sample includes men aged 16 to 54. The sample size is unweighted; all other estimates use the population weights. Predicted education is defined in the text. Central Mexico includes the states of Aguascalientes, Colima, Durango, Guanajuato, Hidalgo, Jalisco, Michoacán de Ocampo, Morelos, Nayarit, Puebla, Querétaro de Arteaga, San Luis Potosí, Sinaloa, and Tlaxcala. The northern boarder states include Baja California, Baja California Sur, Coahuila de Zaragoza, Chihuahua, Nuevo León Sonoora, and Tamaulipas. The Southern States include Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz-Llave, and Yucatán.

Panel B: Differences by age

Age group	Sample size	Migration rate	Fraction of all migrants	Average predicted education of non-migrants	Average predicted education of migrants	Difference
16 to 25	994,644	6.6%	59.0%	8.68	8.00	0.68
26 to 35	736,353	3.9%	26.4%	8.93	8.23	0.69
36 to 45	549,077	2.2%	11.0%	8.34	7.06	1.27
46 to 54	321,493	1.2%	3.6%	6.74	5.48	1.26
Total	2,601,567	4.2%	100.0%	8.43	7.87	0.56

Notes: Sample includes men aged 16 to 54. The sample size is unweighted; all other estimates use the population weights. Predicted education is defined in the text.

Table 6 continued
Panel C: Differences by town size

Town size	Sample size	Migration rate	Fraction of all migrants	Average predicted education of non-migrants	Average predicted education of migrants	Difference
Less than 2500	991,256	7.8%	42.5%	5.70	6.30	-0.60
2500 to 14,999	446,337	5.9%	18.0%	7.41	7.73	-0.32
15,000 to 99,999	278,658	4.2%	13.6%	8.53	8.63	-0.10
100,000 to 499,999	400,862	2.1%	11.2%	9.81	10.06	-0.25
500,000 or more	484,454	2.2%	14.7%	9.89	10.19	-0.30
Total	2,601,567	4.2%	100.0%	8.43	7.87	0.56

Notes: Sample includes men aged 16 to 54. The sample size is unweighted; all other estimates use the population weights. Predicted education is defined in the text.

Table 7: Regression Estimates of Migration Propensity by Region

	Central Mexico			Southern states			Northern boarder states			Mexico City and state		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
County rate of return to schooling	0.484**	0.331*	--	0.046	-0.006		-0.100	-0.260	--	-0.657***	-0.643***	--
	(0.231)	(0.200)		(0.067)	(0.064)		(0.246)	(0.254)		(0.252)	(0.247)	
Years of predicted education	-0.001	0.001	0.005***	0.006***	0.002*	0.006***	-0.003	-0.003	0.000	-0.005*	-0.005*	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Rate of return * years of predicted education	-0.109***	-0.110***	-0.076***	-0.070***	-0.033***	-0.053***	0.012	0.021	0.049**	0.054**	0.059**	0.036**
	(0.029)	(0.025)	(0.025)	(0.013)	(0.011)	(0.010)	(0.023)	(0.026)	(0.022)	(0.024)	(0.023)	(0.016)
Age indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State indicators	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
County indicators	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Sample size	982,906	982,906	982,906	691,429	691,429	691,429	405,462	405,462	405,462	511,081	511,081	511,081
R-squared	0.029	0.038	0.078	0.015	0.025	0.063	0.005	0.007	0.027	0.006	0.006	0.023

Note: Each column is a separate regression, as described in the text. Standard errors in parentheses are adjusted for clustering at the county level. All models use population weights. Sample includes all Mexican men aged 16 to 54. Regions are given in the note to Panel A in Table 6. Asterisks indicate significance at the *** 1%, ** 5%, and * 10% level.

Table 8: Predicted Migration Propensities by Region, Predicted Education, and Return to Education

	Central Mexico		Southern states		Northern boarder states		Mexico City and state	
Migration rate	7.4%		3.0%		2.1%		2.2%	
Fraction of total migrants	63.5%		15.1%		9.3%		12.1%	
Average return to education	0.074		0.087		0.084		0.100	
	Model Specification							
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
State indicators	No	Yes	No	Yes	No	Yes	No	Yes
	Predicted migration rate with return to education = 0.06							
Predicted education:								
6 years	11.9%	11.7%	4.8%	4.6%	3.6%	3.6%	4.2%	0.6%
10 years	8.7%	9.3%	5.4%	4.7%	2.9%	3.1%	3.5%	1.8%
	Predicted migration rate with return to education = 0.10							
Predicted education:								
6 years	11.3%	10.4%	3.3%	3.8%	3.5%	3.0%	3.1%	1.5%
10 years	6.3%	6.2%	2.8%	3.3%	3.0%	2.9%	3.3%	3.3%

Note: The predicted migration rates are calculated from the parameter estimates in Table 7 for a 25 year old person.

Figure 1a: Comparison of Male Populations Counts in the 2000 Mexican and U.S. Censuses

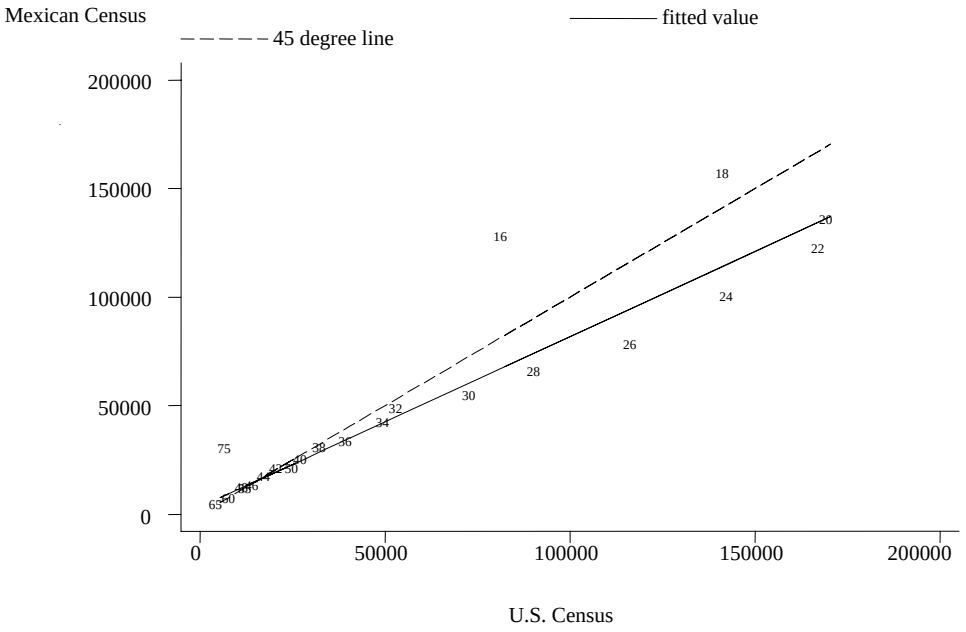
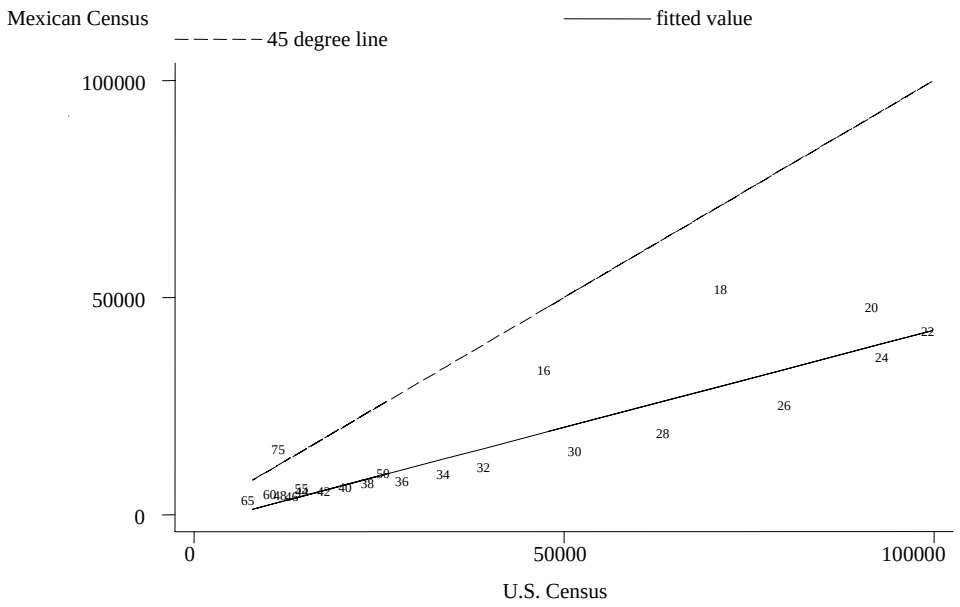


Figure 1b: Comparison of Female Populations Counts in the 2000 Mexican and U.S. Censuses



Note: data points are marked by the minimum of the respective age group. See text for details.