

NBER WORKING PAPER SERIES

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

Pablo Ibarrraran
Darren Lubotsky

Working Paper 11456
<http://www.nber.org/papers/w11456>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2005

We are grateful to Cordelia Reimers, our discussant, for her useful feedback. We also thank seminar participants at Columbia University, the Federal Reserve Bank of Chicago, Indiana University, the NBER, the University of Illinois, the University of Kentucky, and Gadi Barlevy, Ilana Redstone Akresh, Richard Akresh, Todd Elder, Kevin Hallock, Roger Koenker, and Craig Olson for their helpful comments and suggestions. We are naturally responsible for any remaining errors. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2005 by Pablo Ibarrraran and Darren Lubotsky. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census
Pablo Iarraran and Darren Lubotsky
NBER Working Paper No. 11456
June 2005
JEL No. J6, F2

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.

Pablo Iarraran
Office of Evaluation and Oversight
Inter-American Development Bank
1300 New York Avenue NW
Washington, DC 20577
STOP B-0750
pibarraran@iadb.org

Darren Lubotsky
Department of Economics
and Institute of Labor and Industrial Relations
University of Illinois at Urbana-Champaign
504 E Armory Avenue
Champaign, IL 61820
lubotsky@uiuc.edu

I. Introduction

We use data from the 2000 Mexican and United States Censuses 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 have a greater incentive 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. Instead, 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 skill distribution 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, through 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 reduce the relative scarcity of high-skilled labor in Mexico and reduce earnings disparities between high and low-skilled workers. Inequality across Mexican families will be further reduced by remittance income from abroad. Moreover, if economic development and rising educational attainment in Mexico is accompanied by a reduction in the return to skills, then over time there may be a reduction in the size of migrant flows from Mexico to the United States and an increase in the skill composition of future Mexican migrants. 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.¹

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 U.S. 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.² 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.

¹ McKenzie and Rapoport (2004) find that migration tends to reduce inequality within rural Mexican communities.

² For example, Trejo (1997) studies the earnings of Mexicans in the U.S. labor market.

Our main finding is that low-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 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 an 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.³ Let the log earnings of individual i who lives in Mexican county c be given by

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

³ This is a single- index model of skill, similar to that in Chiquiar and Hanson (2005). Borjas (1987, 1991, and 1999) presents a two-index model which allows the rank ordering of workers by skill to be different across countries. All of these models ignore the possibility of back-and-forth migration between Mexican and the United States.

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 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

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

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 prediction 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.

The predictions about migrant selectivity are driven by wage differences between Mexico and the United States that result from differences in the return to skill across countries. These predictions may not hold if time-equivalent migration costs tend to be lower for high skilled Mexicans, as suggested by Chiquiar and Hanson (2005). 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. The presence of these factors may lead migrants to be positively selected even if the wage gain is relatively larger for low-skilled Mexicans. But there are also reasons to believe migration costs may be higher for better skilled workers. For example, high skilled workers may require legalized status to practice their profession in the United States, or they require an extended stay in the United States to acquire U.S. or firm-specific skills. In any event, little is known about the source or magnitude of migration costs.

Though the model captures the essential idea behind wage differences as a driving force behind migration incentives, it contains a number of simplifications that may influence the interpretation of our results. Perhaps most importantly, the rate of return to education in a Mexican county is not necessarily exogenous to the migration process, 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. Finally, recent work stresses the importance of networks and social capital in the migration process.⁴ 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. Our paper does not address the role of these factors in influencing migrant selectivity.

Though the literature on Mexican immigration is vast, there is very little that focuses on the selectivity of migration. Chiquiar and Hanson (2005) compare Mexicans in the 1990 and 2000 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

⁴ For example, see Durand, Massey, and Zenteno (2001); de Janvry, Sadoulet, and Winters (2001); Massey and Singer (1998); and Munshi (2003).

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. Since this ideal dataset does not exist, past researchers have relied on the alternative data sources described in Section II. We take a new approach and use the 2000 Mexican Census to compare the characteristics of Mexican migrants and non-migrants. In doing so we lay out the potential problems and biases associated with the both censuses.

The Mexican Census was conducted in February 2000 by the Instituto Nacional de Estadística Geografía e Informática (INEGI), the Mexican statistical agency. 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.⁵ The data consist of a 10 percent 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.

⁵ 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.

Although the Mexican Census allows us to shed light on some of the limitations of other data sources, the data also have important limitations relative to our ideal dataset: first, 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 do not know migrants' relationship to the household members in Mexico. Second, we do not have any information about households in which all members moved to the United States. 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 compare migrants and non-migrants using the same data source and thus avoid complications stemming from comparing educational attainment measured in the U.S. Census with attainment measured from a different question in the Mexican Census. Second, 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. Third, 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 and Bean *et al* 2001). Costanzo *et al* (2001) suggest that the undercount rate appears to be smaller in the 2000 U.S. Census than it was in the 1990 Census. Clearly, neither the Mexican nor U.S. Censuses provide a fully representative sample of all recent Mexican migrants and they probably provide samples with different sources of bias compared to the universe of all Mexican migrants.

Non-response to Census questions in both data sources also poses a problem for comparing the migrant populations. The U.S. Census Bureau allocates responses for missing values in most cases by imputing a valid response from another respondent in the data. The characteristics used to match "donor" responses to the missing values depend on the particular variable being allocated, but typical characteristics are age, gender, and race. It does not appear that migration status is used in the allocation procedure, so imputed values for migrants could be

coming from American-born respondents.⁶ Of all people recorded in the U.S. Census as being born in Mexico, approximately 13.4 percent have allocated data for their country of birth; 23.5 percent have an allocated year of arrival to the U.S., 18.9 percent have allocated education; and 9.9 percent have an allocated age. The Mexican Census does not allocate data for missing values and 2.3 percent of migrants are missing a value for their age. As we note below, in some cases our conclusions depend on how we handle missing values in both Censuses.

In addition to the sources of discrepancy identified above between the U.S. and Mexican Census counts, and between the Censuses and the universe of all Mexican migrants in the U.S., there are several other sources of discrepancy in coverage: First, 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. Second, 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. This source of discrepancy could of course be larger if migration to the United States was larger than average during February and March of 2000. Third, a back-and-forth migrant could be listed as both someone in the Mexican Census who returned to Mexico and also in the U.S. Census as a current household member. Without knowing the size of this group, it is not clear whether the focus should be on all migrants identified in the Mexican Census or only those who are reported not to have returned to Mexico. Although this distinction is important for

⁶ Hirsch and Schumacher (2004) discuss biases that result from using allocated data in wage regressions. Crease, Ramirez, and Spencer (2001) discuss the quality of the country of birth and Hispanic ethnicity variables in the 2000 Census.

assessing the overall level of coverage in the Mexican Census, it turns out not to be important for our conclusions regarding migrant selectivity.

To shed light on the relative coverage of recent Mexican immigrants enumerated in the 2000 U.S. and Mexican Censuses, we begin in Table 1 and Figures 1 and 2 with a comparison of estimated population counts of Mexican migrants in the Mexican and U.S. Censuses. Panel A of Table 1 shows estimates of the migrant population taken from the Mexican Census. There are 137,910 male migrants aged 16 or older and 38,538 female migrants. The average age of the migrants is about 27 years old for both genders. Using the household weights provided by the Mexican Census, these observations correspond to population estimates of 1,111,895 males and 342,795 females. About 15 percent of Mexican migrants are reported to have returned to Mexico by February 2000. Excluding these individuals leaves 115,760 male migrants aged 16 or older and 33,516 females, corresponding to population estimates of 925,587 males and 296,011 females. These population estimates include migrants with missing values for age.

In panel B we show analogous estimates of the Mexican immigrant population in the United States from the 5% sample of the 2000 United States Public Use Microdata Sample. 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. Using the person weights provided in the census, these sample counts correspond 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 total male and female migrant populations in the Mexican Census are about 84 and 39 percent of the size of the populations in the U.S. Census. Excluding return migrants, the populations in the Mexican Census are 70 and 34 percent of the size of the populations in the U.S. Census. These tabulations include respondents in the U.S. Census with allocated data for country of birth, year of migration, age, or education.

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. Excluding return migrants, the population estimates from the Mexican Census correspond to 90 and 64 percent of these population estimates.

These aggregate population comparisons hide important differences in coverage between the Mexican and U.S. Censuses across age groups. Figure 1 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, five-year age groups from 50 to 70, and men over 70. Figure 2 is the analogous plot for women.⁷ The dashed 45 degree line represents an equal population estimate in the two data sources. The solid line shows the average coverage rate of 90%.⁸ These comparisons are summarized in Table 2. These tabulations exclude respondents in the U.S. Census with allocated age data and also exclude respondents in the Mexican Census with missing age data.

Figure 1 and Table 2 show that young migrant men are actually undersampled in the U.S. Census, in contrast to the pattern for older migrants. There are 36 percent more migrant men aged 16 to 19 in the Mexican Census than in the U.S. Census. Men aged 20 to 31 are underrepresented in the Mexican Census by 77 percent relative to the U.S. Census. In fact, the two data sets disagree over which age group comprises the largest segment of the Mexican migrant population: According to the U.S. Census it is those aged 20 to 21, with those aged 22 to 23 a close second. But according to the Mexican Census the largest group is those 18 to 19. Men aged 32 and older are also underrepresented in the Mexican Census and the degree of under-representation tends to rise with age. The undercount of 16 to 19 year old Mexican migrants in the U.S. Census is likely

⁷ The population estimates of children under age 16 are smaller in the Mexican Census than in the U.S. Census, almost certainly because most children move only when the whole household moves and because of births to Mexicans that occur while in the United States. We thus exclude children from our population comparisons.

⁸ This average coverage rate of 90% in Figure 1 and 40% in Figure 2 are higher than the coverage rates of 84% and 39% reported in Table 1 because the data in Table 1 include respondents with missing or allocated age data, while the data underlying Figures 1 and 2 do not.

caused by the fact they are more likely than older migrants to be in the U.S. illegally and less likely to have set up permanent roots in the United States. For example, we examined the likelihood of being in the U.S. illegally using data from the migration module of the 2002 National Employment Survey and find that about 86 percent of migrants aged 16 to 19 are in the U.S. illegally, compared to 78 percent among migrants aged 20 to 54.⁹

Although the coverage rate for women as a whole is lower than that of men, Figure 2 shows younger women have higher than average coverage compared to older migrants. The lower average coverage rate among women is probably a result of a large number of women only migrating as part of a whole household, and thus not being enumerated at all in the Mexican Census.

The relative undersample of young migrants in the U.S. Census is likely to lead users of that data to overstate the age and skill level of Mexican migrants. To gauge the magnitude of these differences, the right-hand column in Table 2 shows how high school graduation rates of Mexican immigrants in the U.S. Census vary by age. The overall high school graduation rate of Mexicans in the U.S. Census is 24.5 percent, but is only 15.0 percent among migrants aged 16 to 19. When 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 one percentage points to 23.5 percent. In an unreported tabulations we also find that the average annual wage of employed Mexican migrants in the U.S. Census falls by about eight percent when we reweight migrants in different age groups.

To summarize, 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. Both the United States and Mexican Censuses understate the size of the Mexican migration flow, but they have different

⁹ Like the Mexican Census the National Employment Survey asks household members in Mexico whether any other members have recently moved to the United States. Migrants' legal status is reported by the household respondent in Mexico.

shortcomings. The U.S. Census tends to have a greater undersample of migrants aged 16 to 19, who make up about a quarter of all migrants and tend to be less educated than older migrants. The Mexican Census 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. Below we discuss how the relative skills of these unenumerated migrants may impact our conclusions about migrant selectivity. Finally, the last column of Table 2 shows the migration rate of different age groups in Mexico. Since the migration rate is below one percent for Mexicans aged 55 and older, and such migrants make up only two percent of all migrants, we focus the remainder of our analysis on migrants aged 16 to 54.

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 the Mexican Census to migrants in the 2000 U.S. Census. 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. The Mexican Census has a degree-based question and individual degrees (such as primary and secondary) are converted by INEGI into a variable measuring the number of years of schooling. Table 3 shows the distribution of education among non-migrant men sampled in the Mexican Census and migrant men sampled in the U.S. Census. Column 1 shows that 44.8 percent of Mexican men aged 16 to 54 have eight or

fewer years of schooling; 21.9 percent have nine years of schooling; and 25.4 percent have a high school degree or more education. The next four columns show the distribution of education by age and indicate that younger generations are more likely to get a secondary or high school degree than are people aged 35 or older.

The right side of Table 3 shows the distribution of educational attainment among recent Mexican immigrant men in the 2000 U.S. Census. We restrict our sample in this table to those without allocated place of birth, year of arrival, age, or education; this excludes 30 percent of those who would otherwise appear in this table. The U.S. Census also has a grade and degree-based question but naturally the categories are different than in the Mexican Census. 45.4 percent of Mexican migrant men report they have completed eighth grade or fewer years of schooling; 13.6 percent report to have completed ninth grade; 16.4 percent report they completed 10th through 12th grade and do not have a high school degree; and 32.6 percent have a high school degree or more education.¹⁰

These tabulations indicate that Mexican migrants in the U.S. Census come from the upper-middle of the Mexican educational distribution, which echoes the findings of Chiquiar and Hanson (2005). 45 percent of both the U.S. and Mexico samples have between zero and eight years of education. Non-migrant Mexicans are more likely than Mexicans in the U.S. Census to have nine years of education (a secondary school degree), while migrants are more likely to have between 10 years of education and a high school degree. Interestingly, non-migrants are more likely than migrants to have 13 or more years of education. In unreported tabulations we also find that non-migrants are more likely than migrants to have a college degree or more education.

Attempting to credibly compare educational attainment in the U.S. and Mexican Censuses raises several important concerns. First, migrants in the U.S. Census may tend to over-report their education, possibly due to a mistranslation or misunderstanding of the grade and

¹⁰ Unlike the Mexican Census, the United States Census has a category for someone who completed 12 years of schooling but does not have a high school degree. 7.8 percent of Mexican migrants are in this category, which is nearly half the number of people who report to have a high school degree.

degree choices in the U.S. Census.¹¹ 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 9.0 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 16 and 17 year olds. In both countries a person would typically be in their third and final years of high school at ages 16 and 17.

A second potential problem is that the migrants in the U.S. Census are a nonrandom subsample of all migrants. We have already detailed in the previous section 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 the U.S. Census significantly undercounts illegal and low-skilled migrants of all ages. Finally, 30 percent of the migrants in the U.S. Census did not give valid responses to key variables, such as place of birth, year of migration, age, and education. The Census Bureau provides imputed values for all missing data and the values imputed for migrants' education tend to be higher than the average actual reported values. For example, the fraction of Mexican migrants with a high school degree or more education rises from 24.6 percent to 27.1 when individuals with allocated data are included. This increase may result from the Census Bureau using higher educated native-born American respondents to impute education to Mexican immigrants. Thus researchers are faced with a choice of using imputed values that are potentially too large or dropping individuals with imputed values and using a sample with an unknown selection bias.

To alleviate some of the difficulties in comparing Mexicans in two different national censuses, with different sampling schemes and different questions, we next turn to an analysis of educational attainment using only the Mexican Census. We begin in Table 4 with a comparison of

¹¹ For example, a high school degree in Mexico is sometimes referred to as a *bachillerato*, while a bachelor's degree in the United States signifies college completion. Mexicans filling out the U.S. Census may also indicate they have a high school degree when they in fact have a secondary school degree in Mexico, which requires nine years of schooling.

the educational attainment of the highest educated non-migrant in households that contain at least one migrant to the highest educated member of non-migrant households. Migrants themselves are not included in this tabulation because we do not observe their level of education. Higher education among non-migrant family members is associated with higher family income and is likely associated with higher education among migrant members of the same family. If migrants' family members tend to be better educated than non-migrant Mexicans, one might have more confidence in the evidence of positive selection of migrants themselves presented above. But the tabulations in left-hand columns of Table 4 do not bear this out: Members of migrant families are more likely than non-migrant families to have nine years of education or less, 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.

Although these tabulations suggest migrants come from less educated households in Mexico, there are two important problems. 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. The right-hand columns in Table 4 compare the educational attainment of the highest educated woman in non-migrant households, in migrant households, and in migrant households where all migrants are men. The highest educated woman in 55 percent of migrant families has eight or fewer years of education, while only 47 percent of non-migrant families fall in that range.

Women in non-migrant families are more likely than their counterparts in migrant families to have 12 or more years of education. These conclusions are not altered when we restrict the sample of migrant households to just those with male migrants. In unreported tabulations we also find similar conclusions when we restrict attention to women aged 16 to 35, so the higher educational attainment among non-migrant families is not driven by higher education solely among children. In sum, our comparison of educational attainment among non-migrants in Mexico indicates that migrants tend to come from households with lower educated members.

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 migrant and non-migrant Mexican based on their household characteristics and location. We then compare the predicted education of migrants to the predicted education of 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:

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

where the δ s 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. Note also that equation 4 is only estimated on non-migrants who live in non-migrant households because the educational attainment of the non-migrant family members of migrants may be impacted by remittances from migrants living abroad (see Hanson and Woodruff 2003).

Another way to view this procedure is that we are assigning to migrants the average educational attainment of non-migrants who live in towns of the same size in their county, who are the same age, and have a similar family structure. If there are systematic differences between migrants and non-migrants in these narrow cells, we may over- or under-predict migrant education. For example, the theory in Section 2 predicted that migrants will tend to be less educated than non-migrants from across Mexico as a whole. If this prediction holds even within narrow geography, age, and family structure cells, then we are likely overstating the education of migrants. Similarly, if migrants were more likely to work when young and thus attended fewer class sessions, then we are again likely to be overstating the relative skills of migrants. Although we cannot directly test the identifying assumption underlying this procedure that there no

unobserved differences between migrants and non-migrants within the narrow geography, age, and family structure cells, we provide suggestive evidence below that it may be reasonable.¹²

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 also 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 standard deviation of predicted education is about 54 percent of the standard deviation 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 0.57 years less than the average actual education, which implies that unmeasured determinants of schooling exert a positive, though very small, effect on this group. We take this as suggestive evidence that the identifying assumption underlying our method of comparing predicted education is reasonable.

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.¹³ The bottom panel of Table 5 shows the 25th, median, and 75th percentiles of the distributions of actual and predicted education. The lower predicted education of migrants we see at the mean is also

¹² One might also be concerned that since migrants are more likely to come from rural areas, they may also attend lower-quality schools. Thus simply comparing completed years of schooling may understate the relative skills of migrants.

¹³ Note that another way we could test the assumption underlying our use of predicted education is to construct predicted education for Mexican migrants in the U.S. Census. Unfortunately we cannot do this because we do not have their Mexican household or geographic information.

evident throughout the distribution. The predicted education of the median migrant would put him at the 36th percentile in the distribution of predicted education among non-migrants.

The three panels of Table 6 show differences in predicted education by age, region of Mexico, and by town size. One might worry that our finding of lower predicted education among migrants may be driven by differences in the average age of migrants and non-migrants. The tabulations in Panel A of Table 6 shows evidence of negative selection within four age groups, and the degree of negative selection actually increases with age. Interestingly, there is more negative selection among migrants within age groups than there is among migrants of all ages. Because overall educational attainment is considerably lower among those aged 46 to 54, and very few of them are migrants, including them in an aggregate analysis reduces the overall gap in education between migrants and non-migrants.

Two-thirds of Mexican migrant men aged 16 to 54 originate in one of 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.

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.

The results in this section show Mexican migrants enumerated in the 2000 Mexican Census come from less educated households than non-migrants and also have characteristics associated with being less educated, consistent with the predictions of our theoretical model. This evidence of negative selection is at odds with results from the U.S. Census. However, an important limitation of the Mexican Census is that it does not contain migrants whose whole household moved to the United States. Naturally, the degree to which our results would be affected if we were able to include this group depends on the size of the missing group and on the degree of positive or negative selection among them. Clearly, our general conclusion about negative selection will not change if the educational attainment of non-sampled migrants is similar to the educational attainment of sampled migrants, or fraction of migrants not sampled is very small. Our tabulations in Table 1 indicate that the estimated population of Mexican immigrants in the Mexican Census is about 84 percent of the size of the estimated population in the U.S. Census, corresponding to a 16 percent undercount. To the extent that the U.S. Census undercounts migrants as well, the size of the undercount in the Mexican Census may be larger than indicated by the numbers above. Since the migrants missing from the Mexican Census are those whose whole household moved to the United States, another way to approximate the magnitude of the undercount is to note that there are about 292,000 Mexican men in the U.S. Census who are classified as married with their spouse present in the household. If each of these men were missing from missing from the Mexican Census, it would correspond to a 26 percent undercount of men.

Table 7 investigates the degree to which our results would be impacted by positive selection among migrants not in the Mexican Census. The left-hand column shows alternative rates of undercount among Mexican migrants, ranging from zero percent (i.e. a full random sample of Mexican migrants) to 50 percent (i.e. the Mexican Census contains a random sample of 50 percent of the Mexican migrant population and contains none of the other 50 percent). The next columns of the table correspond to alternative assumptions about the degree of positive or

negative selection among this group. The first column assumes that 100 percent of the missing migrants would have predicted education above 8.8 years, which is the median predicted education among non-migrant Mexicans. This is an unrealistically large degree of positive selection, but it gives a lower bound on how large the undercount would have to be to overturn our results. The remaining columns correspond to 75 percent, 50 percent, 35 percent, and 25 percent of the missing migrants having predicted education above 8.8 years. The column corresponding to 35 percent is significant because it is the same degree of negative selection that we estimate for migrants who are enumerated in the Mexican Census. The entries in the table give the fraction of all Mexican migrants (among both the sampled and missing groups) who would have predicted education above 8.8 years. Hence, a number larger than 50 percent indicates overall positive selection of migrants.¹⁴

The results indicate the undercount would have to be greater than 23 percent of the Mexican migrant population to overturn the degree of negative selection that we found among sampled Mexican migrants, and at this level the predicted education of all non-sampled migrants would have to be greater than the median predicted education of non-migrant men. However, this degree of positive selection is certainly unrealistic. If only 75 percent of non-sampled migrants had predicted education above the median, then the undercount would have to be nearly 40 percent. Finally, if there was no selection among non-sampled migrants relative to non-migrants – which still corresponds to non-sampled migrants being significantly better educated than sampled migrants – then there would still be significant negative selection among all Mexican migrants.

We conclude from these tabulations that although negative selection among sampled migrants may overstate the overall degree of negative selection, the undercount rate would have to be very large and there would have to be a significantly large degree of positive selection among non-sampled migrants for there to in fact be positive selection among Mexican migrants

¹⁴ Specifically, the entries in the table are computed as $(1-\alpha)*34.9\% + \alpha*\beta$, where α is the fraction of Mexican migrants not represented in the Mexican Census and β is the fraction of that group that has predicted education above 8.8 years.

as a whole. It seems likely that the small degree of positive selection found by comparing migrants in the U.S. Census with non-migrants in the Mexican Census is driven by a combination of an undersample of young and lower-skilled migrants and over-reporting of education by Mexican migrants. But clearly more research is needed to definitively reconcile these two data sources.

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

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

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

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

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

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

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.¹⁵ 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. More generally it will lead our OLS estimates to understate the negative interaction between schooling and the local return to education. 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

¹⁵ 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.

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.

We estimate equation 7 separately by region using a linear probability model.¹⁶ The results are shown in Table 8 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 9 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 30th and 85th percentiles and the returns to education correspond to approximately the 30th and 90th 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

¹⁶ Linear probability models were considerably quicker to estimate, particularly when we included county fixed effects. We find essentially similar results using probit models.

to education of 0.10. In this scenario, migration rates are 11.3% and 6.3% for the low and high educated Mexicans.

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 8, 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 conclusion is 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, we find that Mexican migrants in the 2000 United States Census are better educated than migrants in the Mexican Census. Although part of this discrepancy may be caused by an undercount of

Mexican migrants whose whole household moved to the United States and were therefore not sampled in the Mexican Census, part may also be due to an undercount of younger, illegal, and low-skilled Mexicans in the U.S. Census.

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.
- Bean, Frank D., Rodolfo Corona, Rodolfo, Tuirán, Karen A. Woodrow-Lafield, and Jennifer van Hook, 2001, "Circular, Invisible, and Ambiguous Migrants: Components of Difference in Estimates of the Number of Unauthorized Mexican Migrants in the United States," *Demography*, Vol. 38, No. 3 (August), pp. 411-422.
- 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, 2005, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, Vol. 113, No. 2 (April), pp. 239-281.
- 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
- Costanzo, Joe, Cynthia Davis, Caribert Irazi, Daniel Goodkind, and Roberto Ramirez, 2001, "Evaluating Components of International Migration: The Residual Foreign Born," Population Division, U.S. Bureau of the Census, Working Paper Series No. 61 (December).
- Crease, Arthur, Roberto Ramirez, and Gregory Spencer, 2001, "Evaluating Components of International Migration: Quality of Foreign-Born and Hispanic Population Data," United States Census Bureau, Population Division Working Paper #65 (December).
- 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.

Hanson, Gordon and Christopher Woodruff, 2003, "Emigration and Educational Attainment in Mexico," Working paper (April).

Hirsch, Barry T. and Edward J. Schumacher, 2004, "Match Bias in Wage Gap Estimates Due to Earnings Imputation," *Journal of Labor Economics*, Volume 22, No. 3 (July), pp. 689-722.

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

McKenzie, David and Hillel Rapoport, 2004, "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico," BREAD Working Paper No. 063 (April).

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 (3328)	1,111,895 (3220)	342,795 (2174)	1,221,598 (3054)	925,587 (2018)	296,011 (2944)
Fraction of U.S. population estimate	66.0%	83.9%	38.9%	55.4%	69.9%	33.6%
Percent female	23.6 (0.1)			24.2 (0.2)		
Age	26.7 (0.03)	26.6 (0.04)	27.0 (0.08)	26.0 (0.04)	26.0 (0.04)	26.2 (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 (3776)	1,324,762 (4500)	880,594 (4060)	1,492,111 (3149)	1,033,060 (3722)	459,051 (3077)
Percent female	40.0 (0.2)			30.8 (0.2)		
Age	28.7 (0.04)	27.8 (0.05)	30.0 (0.07)	27.2 (0.05)	26.4 (0.05)	29.1 (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 repor

Table 2: Differences in Coverage of Male Mexican Migrants in the U.S. and Mexican Censuses

Data source	Population estimate		Age distribution of migrants		Mexican Census population as a percentage of U.S. Census population	HS graduation rate	Migration rate
	U.S. Census	Mexican Census	U.S. Census	Mexican Census		U.S. Census	Mexican Census
Age group							
16 to 19	206,095	280,036	17.0%	25.8%	135.9%	15.0%	7.2%
20 to 31	705,201	543,085	58.3%	50.0%	77.0%	27.3%	5.7%
32 to 54	267,378	241,715	22.1%	22.2%	90.4%	26.0%	2.2%
55 to 65	19,602	15,878	1.6%	1.5%	81.0%	10.0%	0.7%
Over 65	11,606	5,773	1.0%	0.5%	49.7%	9.2%	0.3%
Age 16 and older	1,209,882	1,086,487	100%	100%	89.8%	24.5%	3.7%
HS graduation rate using age distribution from Mexican Census						23.5%	

Note: Samples exclude individuals with missing or allocated age data. The first two columns give population estimates computed as the sum of population weights in each census. The sixth column gives the high school graduation rate tabulated from the U.S. Census, excluding individuals with allocated education data. The migration rate is calculated as the estimated population of male migrants divided by the estimated population of non-migrants in the relevant age group. Both figures are computed from the Mexican Census.

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

Years of completed education	Distribution among Non-migrant Mexicans in the 2000 Mexican				Distribution among Mexican Migrants in the 2000 U.S. Census					
	Age group				Age group					
	16 to 54	16 to 17	18 to 25	26 to 35	36 to 54	16 to 54	16 to 17	18 to 25	26 to 35	36 to 54
0 to 4	17.6%	8.5%	11.2%	13.8%	28.5%	12.6%	7.5%	10.5%	12.3%	23.0%
5 to 8	27.2	32.6	25.5	26.0	28.5	32.8	32.2	33.3	31.0	35.3
9	21.9	29.9	26.2	24.4	14.1	13.6	21.5	15.2	12.0	7.9
10 to 12 without degree	7.8	26.3	10.0	5.9	3.0	16.4	29.9	17.7	14.4	10.0
12 with degree	10.8	2.8	13.9	13.3	8.0	16.2	8.3	17.8	17.6	10.7
13 or more	14.6	0.0	13.1	16.6	17.8	8.4	0.7	5.5	12.7	13.2
Unweighted sample size:	2,406,595	210,044	693,078	686,496	816,977	42,372	2,813	20,946	12,638	5,975
Fraction of sample	100.0%	8.3%	28.8%	28.8%	34.1%	100.0%	6.2%	49.6%	30.1%	14.1%

Note: All estimates and the distribution of each sample across age categories use appropriate population weights. 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. Mexicans with allocated data for country of birth, year of arrival in the U.S., age, or education are excluded from the sample. The category of 10 to 12 years without a high school degree includes people in the Mexican Census who have 10 or 11 years of education and people in the U.S. Census who report completing 10th, 11th, or 12th grade but did not receive a high school diploma.

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

Actual years of education	Highest educated non-migrant in household		Highest educated female non-migrant in household		Households with only male migrants
	Non-migrant households	Migrant households	Non-migrant households	Migrant households	
0 to 4	10.7 %	10.5 %	18.5 %	19.6 %	19.7 %
5 to 8	23.1	31.9	28.2	35.1	37.5
9	21.9	24.1	20.0	20.5	21.0
10 to 11	7.8	8.0	6.7	6.3	6.1
12	14.8	11.7	12.8	9.6	8.8
13 or more	21.8	13.9	13.9	9.0	7.0
Average	9.7	8.9	8.3	7.6	7.4
25th percentile	6	6	6	6	6
Median	9	9	9	8	7
75th percentile	12	12	12	9	9
Number of households	2,148,425	137,667	2,014,849	133,025	96,699

Note: A migrant household is a household that contains at least one migrant.

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
0 to 4	17.3%	23.3%	4.5 and under	6.9%	14.2%	7.4%
5 to 8	27.0	31.5	4.5 to 8.5	35.6	48.1	52.0
9	21.9	20.9	8.5 to 9.5	18.4	14.8	14.7
10 to 11	7.9	6.9	9.5 to 11.5	33.0	19.4	20.7
12	11.0	8.1	11.5 to 12.5	4.6	2.5	3.7
13 or more	14.9	9.3	13.5 and up	1.5	1.0	1.5
Mean	8.5	7.5		8.5	7.4	7.9
25th percentile	6	5		7.1	5.7	6.3
Median	9	7		8.9	7.7	7.9
75th percentile	12	9		10.2	9.4	9.6
Sample size	2,276,862	129,733		2,276,862	129,733	134,743

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 Age, Region, and Town Size

Panel A: 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%	58.0%	8.68	8.00	0.68
26 to 35	736,353	3.9%	25.9%	8.93	8.23	0.69
36 to 45	549,077	2.2%	10.9%	8.34	7.06	1.27
46 to 54	321,493	1.2%	0.0%	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.

Panel B: 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 border 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 border states include Baja California, Baja California Sur, Coahuila de Zaragoza, Chihuahua, Nuevo León, Sonora, and Tamaulipas. The Southern States include Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz-Llave, and Yucatán.

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: Estimates of how the Undercount in the Mexican Census Influences Conclusions about Migrant Self-Selection

Fraction of Mexican migrants missing from 2000 Mexican Census	Fraction of missing migrants with predicted education above the median among non-migrant Mexicans				
	100%	75%	50%	35%	25%
	Fraction of all Mexican migrants with predicted education above the median Mexican non-migrant				
0%	34.9 %	34.9 %	34.9 %	34.9 %	34.9 %
5	38.2	36.9	35.7	34.9	34.4
10	41.4	38.9	36.4	34.9	33.9
15	44.7	40.9	37.2	34.9	33.4
20	48.0	43.0	38.0	34.9	33.0
23	49.9	44.2	38.4	34.9	32.7
25	51.2	45.0	38.7	34.9	32.5
30	54.5	47.0	39.5	34.9	32.0
40	61.0	51.0	41.0	34.9	31.0
50	67.5	55.0	42.5	34.9	30.0

Note: The entries in the left column are alternative measures of the fraction of male Mexican migrants who are not enumerated in the 2000 Mexican Census. The columns to the right give the fraction of all male Mexican migrants with predicted education above 8.8 years, which is the median predicted education among non-migrant males, based on alternative assumptions about the predicted education of the missing migrants.

Table 8: Regression Estimates of Migration Propensity by Region

	Central Mexico			Southern states			Northern border 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** (0.231)	0.331* (0.200)	--	0.046 (0.067)	-0.006 (0.064)	--	-0.100 (0.246)	-0.260 (0.254)	--	-0.657*** (0.252)	-0.643*** (0.247)	--
Years of predicted education	-0.001 (0.002)	0.001 (0.002)	0.005*** (0.002)	0.006*** (0.001)	0.002* (0.001)	0.006*** (0.001)	-0.003 (0.002)	-0.003 (0.002)	0.000 (0.002)	-0.005* (0.003)	-0.005* (0.003)	0.001 (0.002)
Rate of return * years of predicted education	-0.109*** (0.029)	-0.110*** (0.025)	-0.076*** (0.025)	-0.070*** (0.013)	-0.033*** (0.011)	-0.053*** (0.010)	0.012 (0.023)	0.021 (0.026)	0.049** (0.022)	0.054** (0.024)	0.059** (0.023)	0.036** (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 B in Table 6. Asterisks indicate significance at the *** 1%, ** 5%, and * 10% level.

Table 9: Predicted Migration Propensities by Region, Predicted Education, and Return to Education

	Central Mexico		Southern states		Northern border 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 8 for a 25 year old person.

Figure 1: Comparison of Male Population Counts in 2000 Mexican and U.S. Censuses by Age

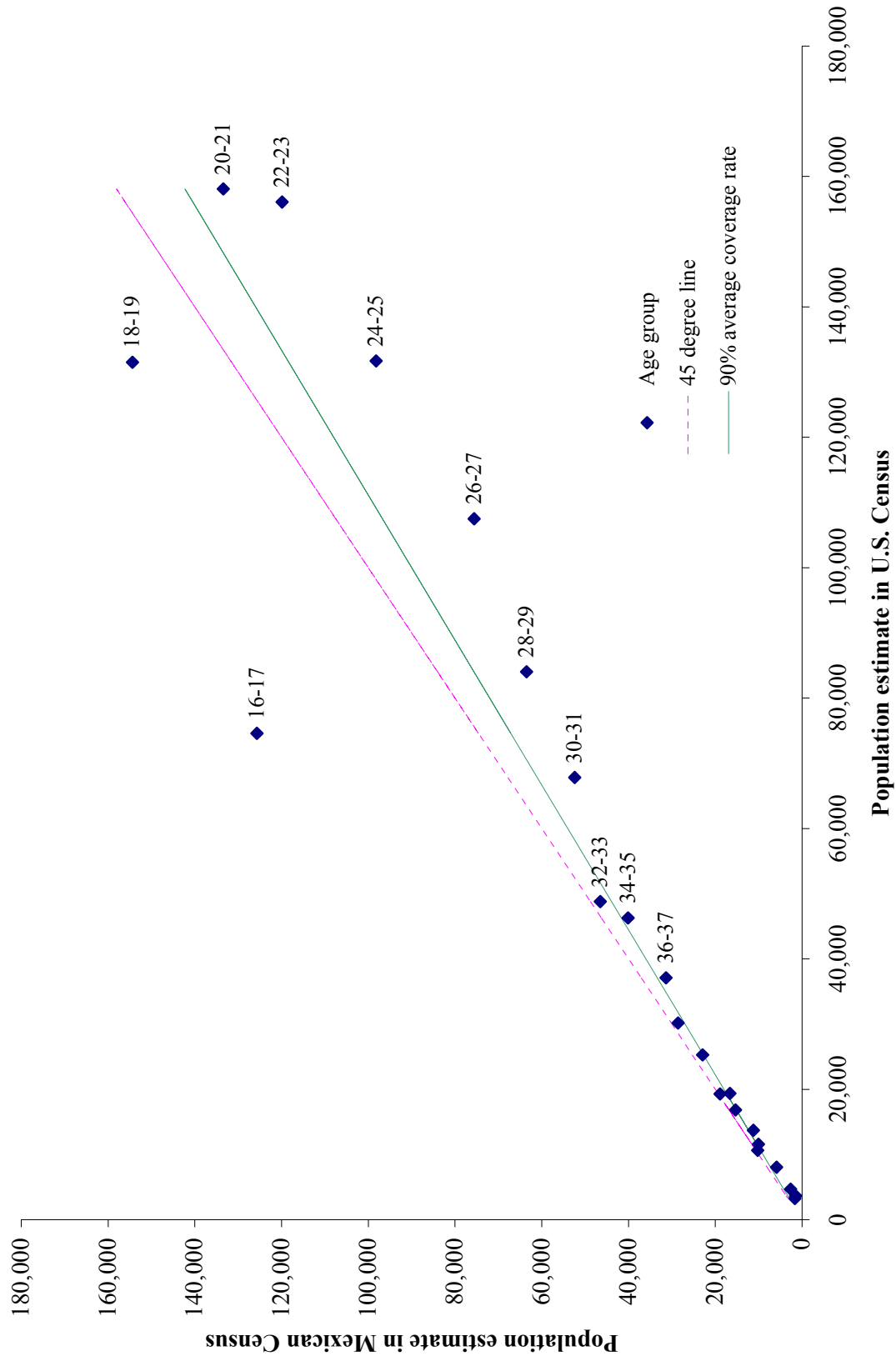


Figure 2: Comparison of Female Population Counts in 2000 Mexican and U.S. Censuses by Age

