Volume Title: Mexican Immigration to the United States

Volume Author/Editor: George J. Borjas, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-06632-0; 978-0-226-06632-5

Volume URL: http://www.nber.org/books/borj06-1

Conference Date: February 11-12, 2005

Publication Date: May 2007

Title: Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census

Author: Pablo Ibarraran, Darren Lubotsky

URL: http://www.nber.org/chapters/c0103

5

Mexican Immigration and Self-Selection
New Evidence from the 2000 Mexican Census

Pablo Ibarraran and Darren Lubotsky

5.1 Introduction

We use data from the 2000 Mexican and U.S. 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 highly 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

Pablo Ibarraran is an Evaluation Economist in the Office of Evaluation and Oversight at the Inter-American Development Bank. Darren Lubotsky is an Assistant Professor of Economics and Labor and Industrial Relations at the University of Illinois at Urbana-Champaign.

For useful feedback, we thank numerous seminar participants as well as Gadi Barlevy, George Borjas, Ilana Redstone Akresh, Richard Akresh, John DiNardo, Todd Elder, Kevin Hallock, Gordon Hanson, Roger Koenker, Justin McCrary, Craig Olson, Cordelia Reimers, and an anonymous referee. We are naturally responsible for any remaining errors.

migrants will be positively selected; that is, they will be more skilled than nonmigrants. 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 important, 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 are 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 preconditions 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 United States 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

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.

immigrant labor market experience in the United States in a wider perspective.

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 overreporting 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 5.3 we describe the 2000 Mexican Census and compare its coverage of migrants with that in the 2000 U.S. Census. Section 5.4 compares the level of education among migrants and nonmigrants. Section 5.5 investigates the relationship between the degree of migrant selection and local returns to education. Section 5.6 concludes.

5.2 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

$$\log(w_{ic}) = \alpha_c + \beta_c S_{ic}, \tag{1a}$$

where S_{ic} is the level of schooling completed by the individual, β_c is the return to schooling in county c, and α_c captures differences in the level of earnings across counties. If the individual were to move to the United States, his log earnings would be determined by

$$\log(w_{iu}) = \alpha_u + \beta_u S_{ic}, \tag{1b}$$

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.

3. This is a single-index model of skill, similar to that in Chiquiar and Hanson (2005). Borjas (1987, 1991, 1999) presents a two-index model that 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 Mexico and the United States.

We assume these rates of return are exogenously given. We also assume that all schooling is completed in Mexico, prior to the migration decision.

A person migrates to the United States if the wage gain plus any nonpecuniary gains outweigh the costs of migration. Denote by C_{ic} the migration costs net of any nonpecuniary 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 cost of migration. The wage gain to individual i were he to move to the United States from county c is given by

$$G_{ic} = (\alpha_u + \beta_u S_{ic}) - (\alpha_c + \beta_{cu} S_{ic}) \tag{2}$$
$$= (\alpha_u - \alpha_c) + S_{ic}(\beta_u - \beta_c).$$

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

$$(\alpha_u - \alpha_c - \pi_{ic}) + S_{ic}(\beta_u - \beta_c) > 0. \tag{3}$$

Equation (3) highlights the important role of differences in the rates of return to education between Mexico and the United States in influencing the types of Mexicans that migrate. Theory and evidence support the notion that the return to schooling acquired in Mexico is considerably higher in Mexico than in the United States. Because education and human capital more generally is a relatively more scarce resource in Mexico than in the United States, it stands to reason that the rate of return is higher in Mexico. Mexicans who acquire their schooling in Mexico and, in particular, in Spanish, may have skills that are not as highly rewarded in, or easily transmittable to, the U.S. labor market. Finally, language barriers may mean that better-educated Mexicans are not able to reap the full benefits of their skills in the U.S. labor market, where English is the predominant language, especially in more highly skilled occupations. While there are a number of empirical challenges in computing comparable rates of return to education for Mexicans in Mexico and in the United States, the difference in the order of magnitude is clear: the coefficient on years of education from an ordinary least squares regression of the log hourly wage on education and a quartic in age is 0.098 in the 2000 Mexican Census and is 0.011 for recent Mexican immigrants in the U.S. Census.[4]

Because the return to education is higher in Mexico than in the United

4. Both estimates use samples of men aged eighteen to fifty-four in the respective censuses. The U.S. data include migrants who arrived in the United States between 1995 and 2000 or who lived in Mexico in April 1995 and do not have allocated data for their place of birth, migration date, schooling, or wage and salary income. A continuous measure of years of schooling is created from the education categories in the census according to the scheme described in Jaeger (1997). Conditioning on a quartic in potential experience instead of age delivers estimates that are slightly larger than those reported in the text.

States, $(\beta_u - \beta_c) < 0$, the wage gain from migrating to the United States is larger for lower-educated Mexicans than it is for higher-educated Mexicans. That is, the relationship between schooling and migrating to the United States should be negative.

Equation (3) also shows that the relationship between schooling and migration should be stronger (i.e., more negative) in areas within Mexico that have relatively larger rates of return to schooling. By contrast, there should be little relationship between schooling and migration in areas with low rates of return. An extreme example would be an area with a rate of return equal to that in the United States, in which case migration and schooling should be unrelated.

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 highly 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 costs 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, highly skilled workers may require legalized status to practice their profession in the United States, or they may 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 important, 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.[5] 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.

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

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 nonmigrant 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 U.S. 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.

A primary source of data on both Mexican residents and migrants to the United States, especially prior to the release of the 2000 Mexican Census, is the Mexican Migration Project. Orrenius and Zavodny (2005) use these 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 nonmigrants. In doing so, we also attempt to shed light on how coverage of Mexican immigrants differs across the two data sources.

5.3 Description of the Mexican Census Data and Its Coverage of Mexican Migrants

With the right data, comparing the skills of migrants to nonmigrants in Mexico is straightforward: the ideal data set 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. Because this ideal data set does not exist, past researchers have relied on the alternative data sources described in section 5.2. We take a new approach and use the 2000 Mexican Census to compare the characteristics of Mexican migrants and nonmigrants. In doing so, we lay out the potential problems and biases associated with 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.[6] A relatively large amount of economic and demographic information was collected about current household members. A much more limited amount of information was 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.[7] The data consist of a 10 percent sample of the Mexican population. Like the U.S. Census, the Mexican Census includes household weights that account for nonresponse. There are 2,312,035 Mexican households in the sample, containing a total of 10,099,182 persons who live in Mexico.

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 data set: 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 nonmigrants 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

6. In Spanish, the census question is "¿Durante los últimos 5 años, esto es, de enero de 1995 a la fecha, alguna persona que vive o vivia con ustedes (en este hogar) se fue a vivir a otro país?" We translate this as "During the last five years, that is, from January 1995 to today, has any person that lives or lived with you (in this household) gone to live in another country?" The instructions for Mexican Census enumerators defines a *household,* according to our translation, as an "Entity formed by one or more individuals, with or without kinship bonds, that regularly reside in the same dwelling and that rely on common consumption of food." The enumerator instructions also make clear that migrants are only counted if they moved abroad directly from the Mexican household. Mexicans that moved from one household to another and then abroad are only included in the migrant roster of the latter Mexican household.

7. 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 do 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.

al. 1998, 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.

Nonresponse 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, race, and, in some cases, Hispanic ethnicity. It does not appear that ethnicity or migration status is used in the allocation procedure for education, so imputed values for migrants could be coming from American-born respondents.[8] 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 in the United States, 18.9 percent have allocated education; and 9.9 percent have an allocated age. The Mexican Census does not include indicators for allocated data, though unlike most variables in the U.S. Census, there are missing values in the data. For example, 2.3 percent of migrants in the Mexican Census 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 previously between the U.S. and Mexican Census counts and between the censuses and the universe of all Mexican migrants in the United States, there are two other sources of discrepancy in coverage. First, the U.S. Census was taken on April 1, 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 flow during these two months is equal to the average flow between 1995 and 2000, this discrepancy will lead to an increase in the U.S. Census count of about one thirtieth or 3.033 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. Second, 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

8. 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.

Census or only those who are reported not to have returned to Mexico. Although this distinction is important for 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 5.1 and figures 5.1 and 5.2 with a comparison of estimated population counts of Mexican migrants in the Mexican and U.S. Censuses. Panel A of table 5.1 shows estimates of the migrant population taken from the Mexican Census. There are 137,910 male migrants aged sixteen or older and 38,538 female migrants of that age. The average age of the migrants is about twenty-

Table 5.1 **Estimates of the Mexican immigrant population in the United States**

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
No. 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
	(3,328)	(3,220)	(2,174)	(3,054)	(2,018)	(2,944)
Fraction of U.S. population estimate (%)	66.0	83.9	38.9	55.4	69.9	33.6
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)

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
No. 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
	(3,776)	(4,500)	(4,060)	(3,149)	(3,722)	(3,077)
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 they came to the United States between 1995 and April 2000 or reported that they lived in Mexico in April 1995. Individuals with missing or allocated age data are included in tabulations.

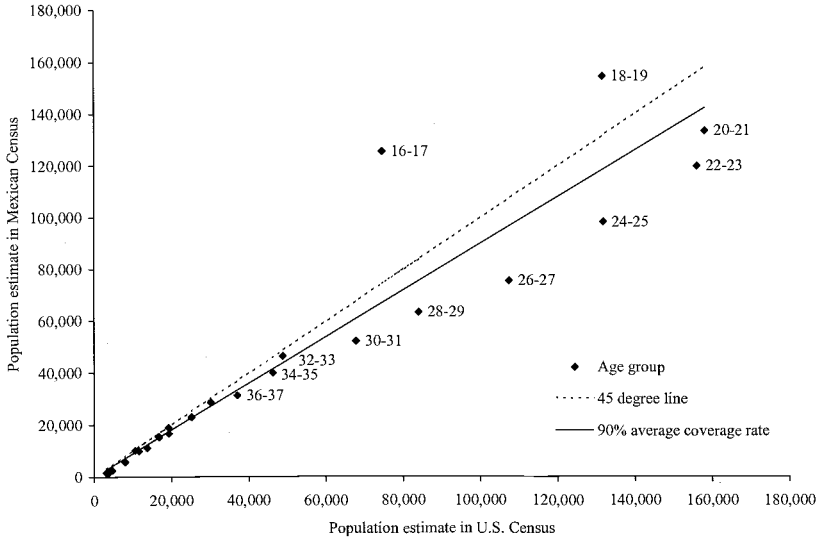

Fig. 5.1 Comparison of male population counts in 2000 Mexican and U.S. Censuses by age

seven 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 and 33,516 female migrants aged sixteen or older, 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 percent sample of the 2000 United States Public Use Microdata Sample. This sample includes all people who report that they came to the United States between 1995 and April 2000 or report that they lived in Mexico in April 1995. There are 62,409 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 twenty-eight years old, and the average female is thirty 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.

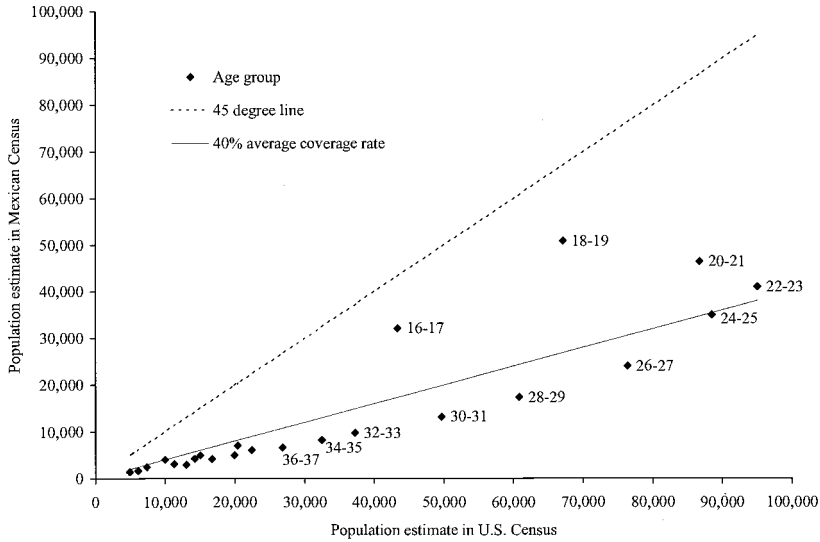

Fig. 5.2 Comparison of female population counts in 2000 Mexican and U.S. Censuses by age

Census that exclude migrants who report themselves as married with spouse present. Because 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 5.1 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 sixteen to fifty, five-year age groups from fifty to seventy, and men over seventy. Figure 5.2 is the analogous plot for women.[9] 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 percent.[10] These comparisons among men are summarized in table 5.2. These tabulations

9. The population estimates of children under age sixteen 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.

10. This average coverage rate of 90 percent in figure 5.1 and 40 percent in figure 5.2 are higher than the coverage rates of 84 percent and 39 percent reported in table 5.1 because the data in table 5.1 include respondents with missing or allocated age data, while the data underlying figures 5.1 and 5.2 do not.

Table 5.2 **Differences in coverage of male Mexican migrants in the United States and Mexican Censuses**

	Population estimate		Age distribution of migrants (%)			HS graduation rate (%)	Migration rate (%)
Data source	U.S. Census	Mexican Census	U.S. Census	Mexican Census	Mexican Census population as a percentage of U.S. Census population	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	100.0	100.0	89.8	24.5	3.7
HS graduation rate using age distribution from Mexican Census =						23.5	

Notes: Samples exclude individuals with missing or allocated age data. Population estimates are computed as the sum of population weights in each census. High school graduation rate is tabulated from the U.S. Census, excluding individuals with allocated education data. Migration rate is calculated as the estimated population of male migrants divided by the estimated population of nonmigrants in the relevant age group. Both figures are computed from the Mexican Census.

exclude respondents in the U.S. Census with allocated age data and also exclude respondents in the Mexican Census with missing age data.

Figure 5.1 and table 5.2 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 sixteen to nineteen in the Mexican Census than in the U.S. Census. Men aged twenty to thirty-one 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 twenty to twenty-one, with those aged twenty-two to twenty-three a close second. But according to the Mexican Census, the largest group is those eighteen to nineteen. Men aged thirty-two and older are also underrepresented in the Mexican Census, and the degree of underrepresentation tends to rise with age. The undercount of sixteen- to nineteen-year-old Mexican migrants in the U.S. Census is likely caused by the fact they are more likely than older migrants to be in the United States illegally and less likely to have established permanent roots in the United States. For example, we examined the likelihood of being in the United States illegally using data from the migration module of the 2002 National Employment Survey and found that about 86 percent of migrants aged sixteen to nineteen are in the United States illegally, compared to 78 percent among migrants aged twenty to fifty-four.[11]

Although the coverage rate for women as a whole is lower than that of men, figure 5.2 shows that 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 in the Mexican Census.

The relative undersample of young migrants in the U.S. Census is likely to lead users of those data to overstate the age and skill level of male Mexican migrants. To gauge the magnitude of these differences, the right-hand column in table 5.2 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 sixteen to nineteen. When we reweight the Mexican immigrants in the U.S. Census to reflect the same distribution across the five age categories as migrants in the Mexican Census, the high school graduation rate falls by 1 percentage point, to 23.5 percent. In unreported tabulations, we also find that the average annual wage of employed Mexican migrants in the U.S. Census falls by about 8 percent when we reweight migrants in different age groups.

11. 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.

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 shortcomings. The U.S. Census tends to have a greater undersample of migrants aged sixteen to nineteen, 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. In the following we discuss how the relative skills of these unenumerated migrants may affect our conclusions about migrant selectivity. Finally, the last column of table 5.2 shows the migration rate of different age groups in Mexico. Since the migration rate is below 1 percent for Mexicans aged fifty-five and older and such migrants make up only 2 percent of all migrants, we focus the remainder of our analysis on migrants aged sixteen to fifty-four.

5.4 Differences in Educational Attainment between Migrants and Nonmigrants

A direct comparison of the educational attainment of migrants and nonmigrants in the Mexican Census is not possible because education of the migrants was not recorded. We instead pursue several alternative strategies: first, we compare educational attainment of nonmigrants 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 nonmigrant Mexicans who live in households that had a migrant to the education of nonmigrants 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 nonmigrants in Mexico.

Most Mexicans have six, nine, twelve, sixteen, or seventeen 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, which range from zero to twenty-two years. Table 5.3 shows the distribution of education among nonmigrant 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 sixteen to fifty-four 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

Table 5.3 **Educational attainment of Mexican-born men in Mexico and the United States, by age group (%)**

	Distribution among nonmigrant Mexicans in the 2000 Mexican Census					Distribution among Mexican migrants in the 2000 U.S. Census				
Years of completed education	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

Notes: 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 United States, age, or education are excluded from the sample. The category of ten to twelve years without a high school degree includes people in the Mexican Census who have ten or eleven years of education and people in the U.S. Census who report completing tenth, eleventh, or twelfth grade but did not receive a high school diploma.

more likely to get a secondary or high school degree than are people aged thirty-five or older.

The right side of table 5.3 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. Over 45 percent of Mexican migrant men report they have completed eighth grade or fewer years of schooling; 13.6 percent report they have completed ninth grade; 16.4 percent report they completed tenth through twelfth grade and do not have a high school degree; and 24.6 percent have a high school degree or more education.[12]

Allocated values of education tend to be higher than the actual reported values of education among Mexican migrants, a problem we suspect may be caused by the use of American-born respondents as "donors" for missing data. In any case, including Mexicans with allocated data in the U.S. Census tabulations tend to raise reported education. For example, in unreported tabulations we find that the fraction of sixteen to fifty-four year olds with zero to eight years of education falls from 45.4 percent to 42.8 percent when the 18,074 sample members with allocated data are included. Including allocated data raises the fraction of Mexicans with 10 or more years of education from 41.0 percent to 44.4 percent.

The tabulations in table 5.3 suggest 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). Forty-five percent of both the U.S. and Mexico samples have between zero and eight years of education. Nonmigrant 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 ten years of education and a high school degree. Interestingly, nonmigrants are more likely than migrants to have thirteen or more years of education. In unreported tabulations, we also find that nonmigrants are more likely than migrants to have a college degree or more education. This general pattern is not altered if we included Mexicans in the U.S. Census who have allocated data.

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 overreport their education, possibly due to a mistranslation or misunderstanding of the grade and degree choices in the U.S.

12. Unlike the Mexican Census, the U.S. Census has a category for someone who completed twelve 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 having a high school degree. The high school degree category in the U.S. Census includes those who passed a high school equivalency exam.

Census.[13] 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 United States may overstate their educational attainment (or understate their age) is that 9.0 percent of sixteen- and seventeen-year-old Mexicans claim to have a high school degree or more education, compared to 3.6 percent of American-born sixteen- and seventeen-year-olds. In both countries a person would typically be in their third and final years of high school at ages sixteen and seventeen.[14]

A second potential problem is that the migrants in the U.S. Census are a nonrandom subsample of all migrants. We have 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. A final problem is the high prevalence of imputed values among Mexican immigrants in the U.S. Census: A full 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 U.S. 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 rises from 24.6 percent to 27.1 when individuals with allocated data are included. This increase may result from the U.S. 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 5.4 with a comparison of the educational attainment of the highest-educated nonmigrant in households that contain at least one migrant to the highest educated member of nonmigrant households. Migrants themselves are not included in this tabulation because we do not observe their level of education. Migrant households are those that had at least one migrant during the past five years, including those in which migrants returned to Mexico. Higher education among

13. 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.

14. Most sixteen- and seventeen-year-old Mexican men in the U.S. Census are the children or relatives of the head of household; less than 3 percent are recorded as the head or spouse. Thus, a parent may be reporting on behalf of sixteen- and seventeen-year-old children. The census does not record which household member filled out the form.

Table 5.4 **Comparison of educational attainment between migrant and nonmigrant households**

Actual years of education	Highest educated nonmigrant in household		Highest educated female nonmigrant in household		
	Nonmigrant	Migrant	Nonmigrant	Migrant	Only male migrants in 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
No. 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.

nonmigrant 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 nonmigrant Mexicans, one might have more confidence in the evidence of positive selection of migrants, presented previously. But the tabulations in the left-hand columns of table 5.4 do not bear this out: members of migrant families are more likely than nonmigrant families to have nine years of education or less, while nonmigrant families are more likely to have twelve or more years of education. Members of nonmigrant families have, on average, about 0.8 years more schooling than those in migrant families.

Although these tabulations suggest that migrants come from less-educated households in Mexico, there are two important problems. First, migrants tend to be men aged sixteen to thirty-five, a group that tends to have high educational attainment within Mexico. Thus, migrant households are likely to be missing their most highly educated members, while nonmigrant households contain them. 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 nonmigrant 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 women across households. Because 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 5.4 compare the educational attainment of the highest educated woman in nonmigrant 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 nonmigrant families fall in that range. Women in nonmigrant families are more likely than their counterparts in migrant families to have twelve or more years of education. These conclusions are not altered when we restrict the sample of migrant households to just those with male migrants, shown in the final column. In unreported tabulations, we also find similar conclusions when we restrict attention to women aged sixteen to thirty-five, so the higher educational attainment among nonmigrant families is not driven by higher education solely among children. In sum, our comparison of educational attainment among nonmigrants in Mexico indicates that migrants tend to come from households with lower-educated members.

Our final and preferred method to compare the relative educational attainment of male migrants and nonmigrants is to generate a predicted level of education for each migrant and nonmigrant male Mexican based on their household characteristics and location. We then compare the predicted education of migrants to the predicted education of nonmigrants.

To predict education, we use an ordered logit framework to model the number of years of schooling, S_{ic}, of individual i who lives in county c as a function of indicator variables for age (A_{ic}), six indicator variables for individuals' town size (T_{ic}), indicator variables for the number of children in the household aged zero to eight (Kid_{1ic}), indicators for the number of children nine to sixteen years old (Kid_{2ic}), indicators for the number of men aged seventeen to thirty-five (Man_{1ic}), indicators for the number of men aged thirty-six and older (Man_{2ic}), indicators for the number of women aged seventeen to thirty-five (Woman_{1ic}), and indicators for the number of women aged thirty-six and older (Woman_{2ic}). Formally, we specify a model for a continuous latent schooling index, S^*_{ic}, and run a separate ordered logit model in each county using all men aged twelve and over who are not migrants and who, furthermore, do not live in a migrant household:

$$(4)\quad S^*_{ic} = \boldsymbol{\delta}_{1c} + \boldsymbol{\delta}_{2c} A_{ic} + \boldsymbol{\delta}_{3c} T_{ic} + \boldsymbol{\delta}_{4c} \text{Kid}_{1ic} + \boldsymbol{\delta}_{5c} \text{Kid}_{2ic} + \boldsymbol{\delta}_{6c} \text{Man}_{1ic} + \boldsymbol{\delta}_{7c} \text{Man}_{2ic} + \boldsymbol{\delta}_{8c} \text{Woman}_{1ic} + \boldsymbol{\delta}_{9c} \text{Woman}_{2ic} + \varepsilon_{ic},$$

where each δ_{kc} is a vector of coefficients that vary by county, and ε_{ic} is the error term. The age indicators 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 an indicator for people over eighty. The town-size indicators correspond to towns with less than 2,500 people; 2,500 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 indica-

tor 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. Equation (4) is only estimated on nonmigrants who live in nonmigrant households because the educational attainment of the migrants' family members may be affected by remittances from migrants living abroad (see Hanson and Woodruff 2003).[15]

Next, we use the coefficient estimates to compute the predicted education for all Mexicans in the data, which includes out-of-sample predictions for nonmigrants who live in migrant households and for migrants themselves. The probability that education is equal to j years is given by $\hat{S}_{ic}(j) = P(S_{ic} = j \mid x_{ic}) = \Lambda(\hat{\alpha}_j - x_{ic}\hat{\delta}_c) - \Lambda(\hat{\alpha}_{j-1} - x_{ic}\hat{\delta}_c)$, where Λ is the logit function, x_{ic} is the set of covariates, and $\hat{\alpha}_j$ is the estimated cut point between schooling level j and $j + 1$.[16] Thus, for each person we have twenty-three probabilities, corresponding to the probability of having zero through twenty-two years of education. We also compute the expected number of years of education, given by

$$\hat{S}_{ic} = E(\hat{S}_{ic} \mid x_{ic}) = \sum_{j=0}^{22} j\hat{S}_{ic}(j). \tag{5}$$

These measures of predicted education can be interpreted as an index of educational attainment or socioeconomic status more generally. The model in equation (4) and the prediction are based only on the individuals' county, age, and household-level characteristics as these are the only variables available for the migrants. Another way to view this procedure is that we are assigning to migrants the average educational attainment of nonmigrants who live in towns of the same size in their county, who are the same age, and who have a similar family structure. If there are systematic unobserved differences between migrants and nonmigrants in these narrow cells, we may over- or underpredict migrant education. For example, the theory in section 5.2 predicted that migrants will tend to be less educated than nonmigrants 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 are no unobserved differences between migrants and nonmigrants within the narrow geography, age, and family

15. An earlier version of this paper used a linear regression to estimate equation (4); (Ibarraran and Lubotsky 2005). The results are nearly identical to those based on the ordered logit model here.

16. The predicted probability of zero years of education is $P(S_{ic} = 0 \mid x_{ic}) = \Lambda(\hat{\alpha}_1 - x_{ic}\hat{\delta}_c)$, and the probability of twenty-two years of education is $P(S_{ic} = 22 \mid x_{ic}) = 1 - \Lambda(\hat{\alpha}_{22} - x_{ic}\hat{\delta}_c)$. More details on the ordered logit model are given in Woolridge (2002).

Table 5.5 **Comparison of actual and predicted education**

	Actual education		Predicted education		
Years of education	Men in nonmigrant households (1)	Nonmigrant men in migrant households (2)	Men in nonmigrant households (3)	Nonmigrant men in migrant households (4)	Migrant men (5)
0 to 4	17.3	23.3	15.4	23.0	17.8
5 to 8	27.0	31.5	31.7	35.1	36.2
9	21.9	20.9	20.9	18.2	19.8
10 to 11	7.9	6.9	7.4	5.7	6.1
12	11.0	8.1	10.1	7.6	8.4
13 or more	14.9	9.3	14.6	10.5	11.7
Mean	8.5	7.5	8.5	7.5	8.0
25th percentile	6	5	7.0	5.7	6.3
Median	9	7	9.0	7.7	8.0
75th percentile	12	9	10.3	9.5	9.7
Sample size	2,276,862	129,733	2,276,862	129,733	134,743

Note: Data include men aged sixteen to fifty-four in the 2000 Mexican Census.

structure cells, we provide suggestive evidence below that it may be reasonable.[17]

Table 5.5 and figure 5.3 compare the distribution of actual and predicted education. For ease of exposition, in table 5.5 the twenty-three levels of education are grouped into six bins. A comparison of actual and predicted education at all levels is shown in figure 5.3. Column (1) of table 5.5 shows the distribution of actual years of education among nonmigrant men aged sixteen to fifty-four who do not live in a migrant household. Column (3) shows the distribution of predicted education for this group. In the table and figure, the predicted frequency of education level j is given by the average predicted probability of education being equal to j. The distribution of actual and predicted education match closely, and there is no systematic pattern in the differences between the two across the levels of education. The mean education of these nonmigrants is eight and a half years and matches the average predicted education, computed according to equation (5).

The model also does well in making out-of-sample predictions of the education of nonmigrants who live in households that also contain migrants. The distribution of actual education, shown in column (2) of table 5.5, closely matches the distribution of predicted education, shown in column (4), for these nonmigrants. The actual and predicted means are both equal to seven and a half years of education. An informal way to assess the ac-

17. One might also be concerned that as 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.

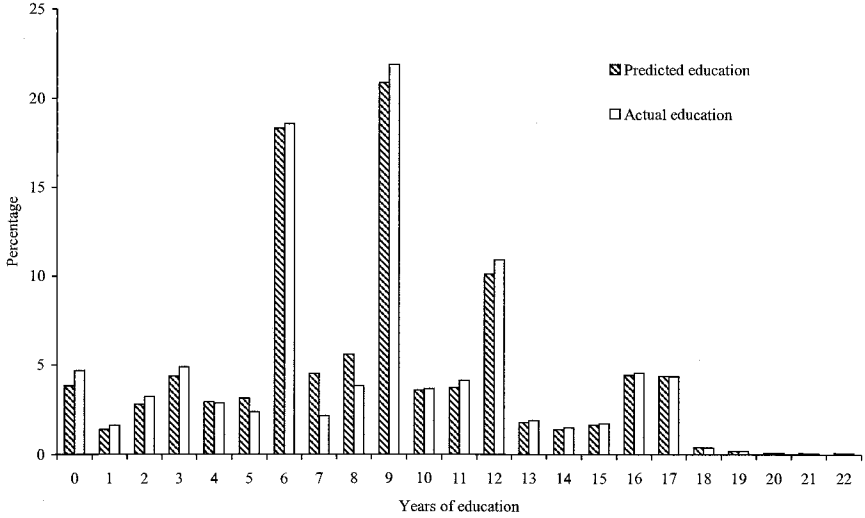

Fig. 5.3 Comparison of predicted and actual education for nonmigrants
Note: Sample excludes nonmigrants who live in households that also contain migrants, as described in the text.

curacy of the model for nonmigrants in migrant households is to note that the root mean square error is 3.57 for the model using nonmigrants who live in nonmigrant households and is 3.66 in the out-of-sample prediction for nonmigrants who live in migrant households.[18] The good fit of the model in predicting education for nonmigrants in migrant households indicates there are not systematic differences in unmeasured determinants of schooling between nonmigrants who live in migrant households and those who do not. We take this as suggestive evidence that the identifying assumption underlying our method of comparing predicted education between nonmigrants and migrants is reasonable.

Column (5) of table 5.5 shows the distribution of predicted education among Mexican migrants. Figures 5.4 and 5.5 compare the distributions of predicted education between migrants and nonmigrants in nonmigrant households and show quite clearly that migrants tend to be less educated than nonmigrants. This is most clearly seen in figure 5.5, which graphs the difference between the height of the distribution of nonmigrants' predicted education and the height of the distribution among migrants. Migrants are more likely than nonmigrants to have between zero and seven years of education. Nonmigrants, by contrast, are more likely to have eight or more years of education. On average, the predicted education of migrants is a

18. The root mean square error is given by $[(1/N)\Sigma_N(S_{ic} - \hat{S}_{ic})^2]^{1/2}$, where i indexes the sample from 1 to N.

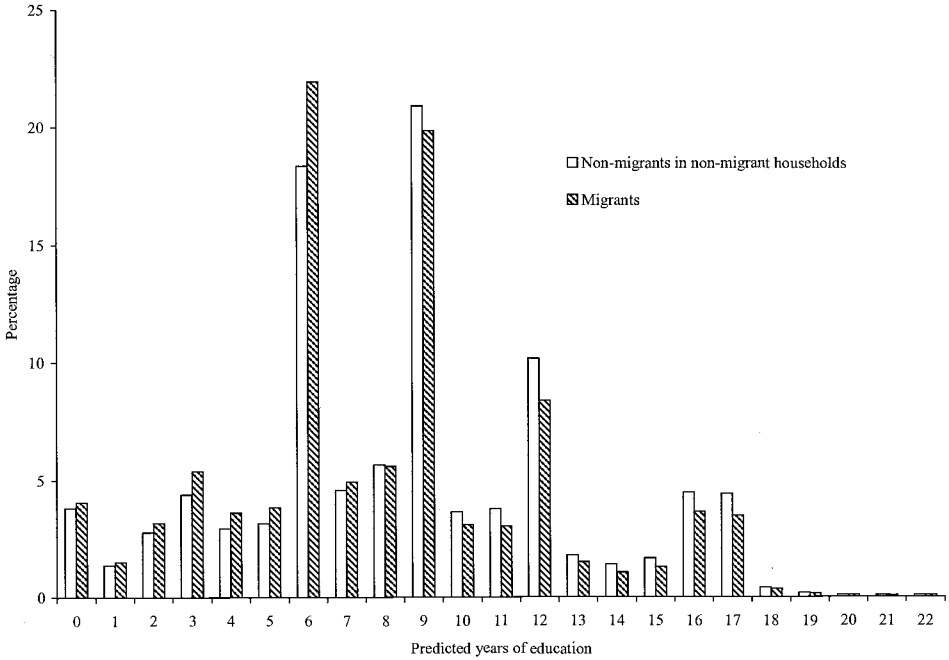

Fig. 5.4 Comparison of predicted education of migrants and nonmigrants
Note: Nonmigrant sample excludes nonmigrants who live in households that also contain migrants, as described in the text.

half year less than the predicted education of nonmigrants who live in nonmigrant households and is a half year more than the predicted education of nonmigrants coming from their same households. If the lack of a discrepancy between actual and predicted education among nonmigrants who live with migrants is any guide, the actual education of migrants is likely to also be less than the actual education of nonmigrants.[19] The bottom panel of table 5.5 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 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 nonmigrants.

The three panels of table 5.6 show differences in predicted education by age, by region within Mexico, and by town size. One might worry that our

19. 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.

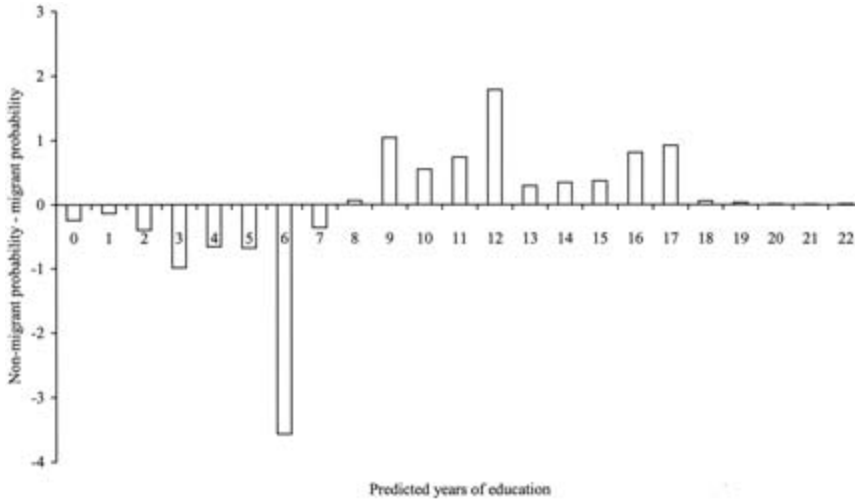

Fig. 5.5 Difference in distribution of predicted education between nonmigrants and migrants

Note: Nonmigrant sample excludes nonmigrants who live in households that also contain migrants, as described in the text.

finding of lower predicted education among migrants may be driven by differences in the average age of migrants and nonmigrants. The tabulations in panel A of table 5.6 show evidence of negative selection within four age groups, and the degree of negative selection is, in fact, larger among older Mexicans than among younger ones. Interestingly, the magnitude of negative selection among migrants within age groups is larger than the negative selection among migrants of all ages. Because overall educational attainment is considerably lower among those aged forty-six to fifty-four and very few of them are migrants, including them in an aggregate analysis reduces the overall gap in education between migrants and nonmigrants.

Panel B shows that two-thirds of Mexican migrant men aged sixteen to fifty-four originate in one of fourteen states in Central Mexico, and the migration rate in this region is 7.5 percent. Interestingly, the nationwide gap in predicted education is entirely driven by the difference in this region. Migrants from the southern states of Mexico are slightly more educated than nonmigrants, while predicted education is approximately equal among migrants and nonmigrants in Mexico City and the state of Mexico and the seven northern border states.

Mexican migrants to the United States tend to come from smaller towns within Mexico. Panel C of table 5.6 shows differences in selection by town size: 42.5 percent of migrants come from towns with populations less than 2,500, but the remaining 57.5 percent of migrants are fairly equally dis-

Table 5.6 **Differences in predicted education between migrants and nonmigrants by age, region, and town size**

	Sample size	Migration rate (%)	Fraction of all migrants (%)	Average predicted education of nonmigrants	Average predicted education of migrants	Difference
			A. Differences by age			
Age group						
16 to 25	985,227	6.7	59.0	8.89	8.20	0.69
26 to 35	720,191	4.0	26.4	8.97	8.24	0.72
36 to 45	529,494	2.3	11.0	8.22	6.89	1.33
46 to 54	306,426	1.3	3.6	6.53	5.27	1.26
Total	2,541,338	4.3	100.0	8.47	7.96	0.51
			B. Differences by region of origin			
Region						
Central Mexico	963,705	7.5	63.5	8.01	7.55	0.46
Southern states	680,742	3.0	15.1	7.41	7.59	–0.18
Northern border states	395,079	2.2	9.3	9.06	9.09	–0.03
Mexico City and state	501,812	2.2	12.1	9.65	9.68	–0.04
Total	2,541,338	4.3	100.0	8.47	7.96	0.51
			C. Differences by town size			
Town size						
Less than 2,500	964,661	8.0	42.5	5.78	6.41	–0.63
2,500 to 14,999	435,627	6.0	18.0	7.43	7.81	–0.38
15,000 to 99,999	273,048	4.3	13.6	8.52	8.70	–0.18
100,000 to 499,999	393,062	2.2	11.2	9.84	10.14	–0.30
500,000 or more	474,940	2.3	14.7	9.92	10.28	–0.37
Total	2,541,338	4.3	100.0	8.47	7.96	0.51

Notes: Sample includes men aged sixteen to fifty-four. 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 Sonoora, and Tamaulipas. The southern states include Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz-Llave, and Yucatán.

tributed among towns with populations of 2,500 and larger. The migration rate is 7.8 percent among all towns with a population of less than 2,500, and the rate tends to fall as town size increases. At the same time, average education of both migrants and nonmigrants tends to rise with town size. Conditional on town size, migrants tend to be more educated than nonmigrants. Negative selection overall is driven by the fact that migrants tend to come from small towns, where educational attainment is very low, while the average nonmigrant lives in a larger city, where educational attainment tends to be higher. These patterns 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 that Mexican migrants enumerated in the 2000 Mexican Census come from less-educated households than nonmigrants 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. 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 nonsampled migrants is similar to the educational attainment of sampled migrants or the fraction of migrants not sampled is very small. Our tabulations in table 5.1 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 preceding numbers. Because 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 spouse present in the household. If each of these men were missing from the Mexican Census, it would correspond to a 26 percent undercount of men.

Table 5.7 investigates the degree to which our results would be affected by positive selection among migrants not enumerated in the Mexican Census. The left-hand column shows alternative hypothetical undercount rates among Mexican migrants, ranging from 0 percent (i.e., the Mexican Census actually contains 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 five columns of the table correspond to alternative assumptions about the degree of positive or negative selection among the nonsampled group. The first column assumes that 100 percent of the missing migrants would have predicted education above 8.9 years, which is the median predicted education among all nonmigrant 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 for there to be negative selection among the enumerated migrants and positive selection overall. The remaining columns correspond to 75 percent, 50 percent, 36 percent, and 25 percent of the missing migrants having predicted education above 8.9

Table 5.7 **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 nonmigrant Mexicans				
	100%	75%	50%	36%	25%
	Fraction of all Mexican migrants with predicted education above the median Mexican nonmigrant				
0	36.0	36.0	36.0	36.0	36.0
5	39.2	38.0	36.7	36.0	35.5
10	42.4	39.9	37.4	36.0	34.9
15	45.6	41.9	38.1	36.0	34.4
20	48.8	43.8	38.8	36.0	33.8
22	50.1	44.6	39.1	36.0	33.6
25	52.0	45.8	39.5	36.0	33.3
30	55.2	47.7	40.2	36.0	32.7
40	61.6	51.6	41.6	36.0	31.6
50	68.0	55.5	43.0	36.0	30.5

Notes: 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 8.9 years (the median predicted education among all nonmigrant men) based on alternative assumptions about the predicted education of the missing migrants.

years. The column corresponding to 36 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 predict education above 8.9 years. Hence, an entry larger than 50 percent indicates overall positive selection of migrants, and an entry smaller than 50 percent indicates overall negative selection.[20]

The results indicate the undercount rate among Mexican migrants would have to be greater than 22 percent to overturn the degree of negative selection that we find among sampled Mexican migrants, and at this undercount rate the predicted education of all nonsampled migrants would have to be greater than the median predicted education of nonmigrant men. However, this degree of positive selection is certainly unrealistic. If only 75 percent of nonsampled migrants had predicted education above the median, then the undercount would have to be nearly 40 percent. Finally, if there was no selection among nonsampled migrants relative to nonmigrants—which still corresponds to nonsampled migrants being significantly better educated than sampled migrants—then there would still be significant negative selection among all Mexican migrants.

20. Specifically, the entries in the table are computed as $(1 - \alpha) \cdot 36\% + \alpha \cdot \beta$, 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 8.9 years.

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 nonsampled migrants for there to, in fact, be positive selection among Mexican migrants as a whole. It seems likely that the small degree of positive selection found by comparing migrants in the U.S. Census with nonmigrants in the Mexican Census is driven by a combination of an undersample of young and lower-skilled migrants and overreporting of education by Mexican migrants. But clearly more research is needed to definitively reconcile these two data sources.

5.5 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_{ic} who lives in Mexican county c is given by

$$G_{ic} = (\alpha_u - \alpha_c) + S_{ic}(\beta_u - \beta_c), \tag{6}$$

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 nonpecuniary gains outweigh the costs of migration. Formally, define the indicator variable M_{ic} to equal one if person i migrates to the United States and zero otherwise. Then $M_{ic} = 1$ if $G_{ic} > \pi_{ic}$, where π_{ic} are time-equivalent migration costs net of any nonpecuniary gains. Alternatively, $M_{ic} = 1$ if

$$(\alpha_u - \alpha_c - \pi_{ic}) + S_{ic}(\beta_u - \beta_c) > 0. \tag{7}$$

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

$$\Pr(M_{ic} = 1) = A_{ic} + d_c + \lambda_1 \hat{S}_{ic} + \lambda_2 \hat{\beta}_c + \lambda_3 \hat{\beta}_c \hat{S}_{ic} + v_{ic}, \tag{8}$$

where A_{ic} is a full set of indicators for each age from sixteen to fifty-four, and d_c is either a set of state or county indicators.[21] $\hat{S}_{ic}$ is the predicted schooling level, computed according to equation (5). $\hat{\beta}_c$ is an estimate of the returns to schooling in county c, as described below.

21. The main effect of the 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 effects.

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 of origin. If $\beta_u - \beta_c < 0$, then according to equation (6), 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 eighteen to fifty-four. We run this model separately by county and weight each observation by the Mexican 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 less-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 estimates of equation (8) toward finding a positive effect of the interaction between migrants' predicted education and their local return to schooling. More generally, it will lead our ordinary least squares 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 ordinary least squares models.

Our estimates of the return to schooling may also be influenced by the lack of earnings data for workers in the informal sector. Missing earnings data may be particularly problematic in rural areas and among 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 affect the relative returns to education across areas.

We estimate equation (8) separately by region using a linear probability model.[22] The results are shown in table 5.8 and, at least for the central and southern regions of Mexico, are 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.[23] 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.

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

23. The standard errors in table 5.8 adjust for clustering at the county level but are not adjusted for the fact that predicted education and the country return to schooling are themselves estimated variables.

Table 5.8 **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	0.520**	0.366*		0.055	0.004		–0.078	–0.236		–0.652***	–0.639**	
	(0.232)	(0.198)		(0.069)	(0.065)		(0.247)	(0.255)		(0.254)	(0.248)	
Years of predicted education	–0.002	0.000	0.006***	0.006***	0.002*	0.006***	–0.003	–0.002	0.001	–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	–0.111***	–0.113***	–0.085***	–0.073***	–0.035***	–0.059***	0.011	0.019	0.042**	0.054**	0.058**	0.035**
	(0.029)	(0.025)	(0.026)	(0.013)	(0.012)	(0.011)	(0.024)	(0.026)	(0.021)	(0.024)	(0.023)	(0.017)
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	963,705	963,705	963,705	670,657	670,657	670,657	395,079	395,079	395,079	501,812	501,812	501,812
R^2	0.030	0.039	0.079	0.015	0.026	0.066	0.005	0.007	0.028	0.006	0.006	0.024

Notes: 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 sixteen to fifty-four. Regions are given in the note to panel B in table 5.6.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

It is difficult to know why the regression results support the predictions of the theoretical model within Central and Southern Mexico but do not support the predictions in the northern states or Mexico City and the state of Mexico. Central Mexico is the source of most migrants, and has been for some time. Migration networks may be developed in this area to the point that migration costs are generally low for most families, and migration decisions largely reflect wage differences. More generally, a potentially interesting avenue for future research is to explore differences in the level and source of migration costs between Mexican regions as well as differences in migration propensities.

To help interpret the magnitude of our regression results, in table 5.9 we show predicted migration probabilities derived from our regressions. These rates refer to twenty-five-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

Table 5.9 **Predicted migration propensities, by region, predicted education, and return to education**

	Central Mexico		Southern states		Northern border states		Mexico City and state	
Migration rate	7.5%		3.0%		2.2%		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	12.2%	11.9%	5.1%	4.6%	3.7%	3.6%	4.7%	4.3%
10 years	8.8%	9.4%	5.6%	4.7%	2.9%	3.1%	3.8%	3.5%
	Predicted migration rate with return to education = 0.10							
Predicted education								
6 years	11.6%	10.7%	3.5%	3.8%	3.6%	3.2%	3.3%	3.2%
10 years	6.4%	6.3%	2.9%	3.3%	3.0%	3.0%	3.3%	3.3%

Note: The predicted migration rates are calculated from the parameter estimates in table 5.8 for a twenty-five-year-old person.

each region. Counties in Central Mexico with a rate of return to education of 0.6 would tend to have a migration rate of 12.2 percent among Mexicans with six years of education, based on our results from model 1. The migration rate falls by 3.4 percentage points to 8.8 percent among those with ten years of education who live in the same area. Areas in Central Mexico with a rate of return to schooling of 0.10 are predicted to have a 5.2 percentage point difference in migration propensities between higher- and lower-educated Mexicans. In this scenario, migration rates are 11.6 percent and 6.4 percent for the lower- and higher-educated Mexicans. Our findings are similar when we use models 2 and 3 from table 5.8 to generate the predicted migration rates.

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 is 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. Using the estimates from specification 1, the migration rate is predicted to be 5.1 percent among those with six years of predicted education and 5.6 percent among those with ten 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 3.5 percent among lower-educated Mexicans and 2.5 percent among higher-educated Mexicans. Echoing our regression results in table 5.8, higher returns to education in the northern border states or in Mexico City and state do not generate an increase in the magnitude of negative selection.

5.6 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 nonmigrants. 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 U.S. 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.

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,* 1–88. Mexican Ministry of Foreign Affairs and U.S. Commission on Immigration Reform. Austin, TX: Morgan Printing.

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* 38 (3): 411–22.

Borjas, George. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 77 (4): 531–53.

———. 1991. Immigration and self-selection. In *Immigration, trade, and the labor market,* ed. John M. Abowd and Richard B. Freeman, 29–76. Chicago: University of Chicago Press.

———. 1999. The economic analysis of immigration. In *Handbook of labor economics.* Vol. 3A, ed. Orley C. Ashenfelter and David Card, 1697–1760. 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* 113 (2): 239–81.

Costanzo, Joe, Cynthia Davis, Caribert Irazi, Daniel Goodkind, and Roberto Ramirez. 2001. Evaluating components of international migration: The residual foreign born. U.S. Census Bureau Population Division Working Paper Series no. 61. Washington, DC: U.S. Census Bureau, December.

Crease, Arthur, Roberto Ramirez, and Gregory Spencer. 2001. Evaluating components of international migration: Quality of foreign-born and Hispanic population data. U.S. Census Bureau Population Division Working Paper Series no. 65. Washington, DC: U.S. Census Bureau, December.

Cuecuecha, Alfredo. 2003. The educational characteristics of the immigrants from Mexico to the U.S. Instituto Technológico Autónomo de México-Centro de Investigacion Economica, Working Paper.

de Janvry, Alain, Elisabeth Sadoulet, and Paul Winters. 2001. Family and community networks in Mexico-U.S. migration. *Journal of Human Resources* 36 (1): 159–84.

Durand, Jorge, Douglas Massey, and Rene Zenteno. 2001. Mexican immigration to the United States: Continuities and changes. *Latin American Research Review* 36 (1): 107–27.

Hanson, Gordon, and Christopher Woodruff. 2003. Emigration and educational attainment in Mexico. University of California at San Diego, Department of Economics, Working Paper.

Hirsch, Barry T., and Edward J. Schumacher. 2004. Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics* 22 (3): 689–722.

Ibarraran, Pablo, and Darren Lubotsky. 2005. Mexican immigration and self-

selection: New evidence from the 2000 Mexican Census. NBER Working Paper no. 11456. Cambridge, MA: National Bureau of Economic Research, June.
Jaeger, David A. 1997. Reconciling the old and new Census Bureau education questions: Recommendations for researchers. *Journal of Business and Economic Statistics* 15 (3): 300–309.
Massey, Douglas, and Audrey Singer. 1998. The social process of undocumented border crossing among Mexican migrants. *International Migration Review* 32: 561–92.
McKenzie, David, and Hillel Rapoport. 2004. Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. Bureau for Research and Economic Analysis of Development Working Paper no. 063. Cambridge, MA: BREAD, April.
Munshi, Kaivan. 2003. Networks in the modern economy: Mexican migrants in the U.S. labor market. *Quarterly Journal of Economics* 118 (2): 549–97.
Orrenius, Pia, and Madeline Zavodny. 2005. Self-selection among undocumented immigrants from Mexico. *Journal of Development Economics* 78 (1): 215–40.
Trejo, Stephen. 1997. Why do Mexican Americans earn low wages? *Journal of Political Economy* 105 (6): 1235–68.
Woolridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.