NBER WORKING PAPER SERIES

SELF-SELECTION AND INTERNATIONAL MIGRATION: NEW EVIDENCE FROM MEXICO

Robert Kaestner
Ofer Malamud

Working Paper 15765
http://www.nber.org/papers/w15765

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2010

Self-Selection and International Migration: New Evidence from Mexico
Robert Kaestner and Ofer Malamud
NBER Working Paper No. 15765
February 2010
JEL No. J24,J61

ABSTRACT

This paper uses data from the Mexican Family Life Survey (MxFLS) to examine the patterns of selection of male, Mexican migrants to the United States. We confirm previous findings that Mexican migrants are selected from the middle of the education distribution, but show that there is no evidence for selection of migrants on cognitive ability. We demonstrate that migrants are also selected from the middle of the observed skill distribution, as measured by predicted wages. However, controlling for proxies of the costs of migration, we find substantially less evidence of "intermediate selection" on observed skill. We find little evidence for selection on unobserved skill, with or without controls for the costs of migration. Finally, we show directly that the decision to migrate is highly correlated with differential returns to observable skill and the costs of migration. Overall, these findings are consistent with the predictions of the canonical model of migration.

Robert Kaestner
Institute of Government and Public Affairs
University of Illinois
815 West Van Buren Street, Suite 525
Chicago, IL 60607
and NBER
kaestner@uic.edu

Ofer Malamud
Harris School of Public Policy Studies
University of Chicago
1155 East 60th Street
Chicago, IL 60637
and NBER
malamud@uchicago.edu

1 Introduction

Mexican immigrants account for a third of all foreign born residents in the United States and are the fastest growing immigrant group. In 2008, it is estimated that 12.7 million Mexican immigrants lived in the U.S. and this figure represents approximately 4 percent of the U.S. population and 11 percent of Mexico's population.[1] Understanding the pattern of self-selection among Mexican immigrants to the United States is relevant for both research and policy. Given the magnitudes involved, selective migration out of Mexico may have large effects on the Mexican labor force. If migrants are negatively selected on skills (i.e., come from the lower end of the skill distribution), then this will tend to reduce the relative scarcity of high skilled labor and reduce earnings disparities between high and low-skilled workers in Mexico. Positive selection of migrants would have opposite effects. The skill-composition of Mexican immigrants in the United States also has important implications for U.S. labor market and immigration policy. Negative selection of Mexican immigrants will tend to reduce the wages of low-skilled native-born persons and exacerbate earnings inequality, although the evidence in support of this hypothesis remains controversial (Borjas, Katz and Freeman, 1996; Lalonde and Topel, 1997; Borjas, 2003; Card, 2005). Finally, studying the pattern of self-selection from Mexico can serve as a useful test of Borjas's (1987) canonical model of migration. Given the documented higher returns to schooling and greater dispersion of wages in Mexico than in the United States, Borjas' theory predicts that Mexican migrants to the United States would be less-skilled than their counterparts who remain in Mexico. However, this prediction has not always been supported by empirical evidence in previous studies.

This paper uses data from the Mexican Family Life Survey (MxFLS) to examine the nature of selection of Mexican migrants to the United States. The MxFLS is ideally suited to study the pattern of selection of migrants. The 2002 baseline survey of the MxFLS contains detailed, individual-level information on education, labor market participation and earnings, past migration experience, migration networks, credit availability, assets, and cognitive ability for a nationally

[1] Fact Sheet: Mexican Immigrants in the United States, 2008; Pew Hispanic Center, http://pewhispanic.org/files/factsheets/47.pdf, website last accessed 2/4/10

representative sample of Mexicans. Most important, there is information about whether individuals migrated to the United States by the time of the second round survey in 2006. Thus, we observe earnings, education, and other components of labor market skills for both migrants and non-migrants prior to their migration from Mexico to the United States. Using the MxFLS enables us to address many of the data limitations that have plagued earlier studies. For example, studies that used the U.S. Census may miss many Mexican migrants because of the undercounting of immigrants, and these studies may overstate the education of migrants if immigrants obtain additional schooling after arrival in the U.S. Other studies that have used Mexican surveys generally miss individuals in households in which all members have migrated to the United States. Furthermore, the MxFLS provides information on aspects of skills such as cognitive ability that are usually unmeasured, as well as a rich set of individual and household characteristics that can help control for other potential costs and benefits associated with the migration decision.

We find, in line with the previous literature, that Mexican migrants are selected from the middle of the education distribution, or what has been referred to as "intermediate" selection. However, there is no evidence for selection on another dimension of skill, namely cognitive ability. We also demonstrate that migrants are selected from the middle of the distribution of observed skills, as measured by predicted wages. In contrast, we find little consistent evidence of selection of migrants on unobserved skill. Our findings are quite different once we control for proxy variables for the costs of migration. In this case, estimates suggest no significant selection of migrants on observed skill, although estimates are somewhat imprecise. The substantial change in the pattern of selection when these controls for the costs of migration are included raises the possibility that analyses with more complete controls would yield evidence of negative selection. With respect to unobserved skill, adding controls for the costs of migration has little effect on estimates, and estimates suggest essentially no selection on unobserved skill.

Finally, we present direct evidence to support the canonical model of migration (Borjas, 1987). We find that Mexican males are more likely to migrate to the U.S. when the return to observed skill is larger in the U.S. than in Mexico in ways consistent with theory. In addition, we show that

proxy variables for the cost of migration are strongly correlated with the probability of migrating and correlated with observed skill; controlling for these costs diminishes, even eliminates, evidence of selection from the middle of the observed skill distribution.

The paper is organized as follows: Section 2 reviews the related literature on the self-selection of Mexican migrants. Section 3 outlines a simple model of migration in which mobility costs are allowed to vary with schooling and with unobserved skills. Section 4 describes the data from the Mexican Family Life Survey and the empirical strategy. Section 5 presents our findings on the selection of migrants by observed and unobserved skills. Section 6 discusses these results in light of previous research, and Section 7 concludes.

2 Related Literature

There are only a few papers that have examined the pattern of selection on skills among Mexican immigrants to the United States. Chiquiar and Hanson (2005) used the 1990 and 2000 Mexican and U.S. Censuses to evaluate the selection of Mexican immigrants in terms of observable skills and found evidence of "intermediate" selection, which is inconsistent with Borjas's model given what is known about the returns to education in Mexico and the U.S.[2] Chiquiar and Hanson (2005) first compared educational attainment of non-migrants in Mexican Census to migrants in the 2000 U.S. Census, and showed that Mexican immigrants in the United States are drawn from the middle of the Mexican education distribution. Next, they compared the (predicted) wage distribution for residents of Mexico with the (predicted) wage distribution of Mexican immigrants in the United States had they been paid according to the skills prices in Mexico. Wages were predicted using age, education, and marital status and used as a proxies for observed skill. As with education, they found that migrants are concentrated in the middle of Mexico's (predicted) observed skill distribution.

The use of the U.S. and Mexican Censuses presents several empirical problems for Chiquiar and

[2]Feliciano (2005) used the Mexican and U.S. Censuses and reported positive selection on education of Mexican migrants. On average, migrants had one to two years more education than non-migrants depending on the year (1960, 1970, 1990, 2000). Feliciano only examined mean differences and not differences in the entire education distribution.

Hanson (2005): the likely undercount of immigrants in the U.S. Census; the possibility that Mexican immigrants can obtain additional schooling after arriving in the United States; return migration; and the likelihood that unobservable characteristics are correlated with education. Results remained relatively unchanged after addressing these potential problems, but despite their comprehensive assessments, these concerns cannot be completely alleviated.

Ibarraran and Lubotsky (2007) used the 2000 Mexican Census to estimate differences in the educational attainment between migrants and non-migrants. They replicated Chiquiar and Hanson's (2005) finding that Mexican immigrants in the 2000 U.S. Census have more education (and are older) than non-migrants in the Mexican Census. However, they further investigated potential over-reporting of education by migrants and undercounting of younger migrants in the U.S. Census. Ibarraran and Lubotsky (2007) exploited the fact that, in the Mexican Census, heads of households were asked to list current or past household members who had lived abroad during the preceding five years. Unfortunately, the heads of household only reported a limited set of individual characteristics: age, gender, Mexican state of origin, and migration patterns. Since no information about educational attainment of migrants was available, the authors used measured characteristics to predict educational attainment. In contrast to Chiquiar and Hanson (2005), they found evidence of negative selection on education; low-educated Mexicans were more likely than high-educated Mexicans to migrate to the United States, and the degree of negative selection was larger in regions with higher returns to schooling. However, this approach is based on the assumption that there are no differences between migrants and non-migrants beyond those measured by the limited set of characteristics used to predict education.

Fernández-Huertas Moraga (2008) used data from the Encuesta Nacional de Empleo Trimestral (ENET) survey to overcome some of the limitations associated with using Census data. The ENET survey follows households for five consecutive quarters and includes reports on whether any household member has left for the United States. Consequently, it is possible to examine the selectivity of migrants on their education attainment and actual earnings prior to migration. Fernández-Huertas Moraga found negative selection on wages and "intermediate to negative selection" on education

for men aged 16 to 65. The ENET is not subject to the problems of undercounting of migrants and over-reporting of migrant educational attainment that are present in U.S. Census. However, this approach misses persons in households in which all members have migrated to the United States.[3] Furthermore, as we discuss later, the inclusion of younger men aged 16 to 20 is potentially problematic given their low rates of labor market participation and lack of information on wages.

Orrenius and Zavodny (2005) used data from the Mexican Migration Project (MMP) to examine how various factors influence the selectivity of undocumented migrants over time.[4] They found that migrants were drawn from the middle of the education distribution. They did not examine other measures of skill. While the MMP includes both migrants and non-migrants living in Mexico, it does not include Mexican-born household heads who have migrated to the U.S. permanently. Furthermore, the data from the MMP are collected retrospectively, which may result in recall bias.

Finally, McKenzie and Rapoport (2007) used data from the 1997 Encuesta Nacional de la Dinámica Demográfica (ENADID) survey to examine selection of Mexican migrants on education.[5] The survey has information on migrants in the U.S. who have a (coresident) family member living in Mexico but misses migrants who have permanently left Mexico and who do not have family member that remained in Mexico. The main contribution of McKenzie and Rapoport (2007) was incorporating the migration history of the community, as a measure of the cost of migration, into the analysis. McKenzie and Rapoport (2007) reported that migrants are negatively selected on education in communities that have a high proportion of persons who migrated, but that selection on education is positive in communities that do not have high rates of migration.

This brief review of the literature highlights the significant data limitations associated with previous research on the pattern of selection of Mexican migrants to the U.S. All the papers reviewed above have developed novel and innovative uses of data to address the question of migrant

[3] As noted by Fernández-Huertas Moraga (2008), individuals moving to the US with their entire household are more likely to have higher education and wages creating a bias towards negative selection.

[4] Munshi (2003) also uses the MMP to show that Mexican migrants are more likely to be employed and hold a higher paying job when their network in the United States is larger.

[5] Durand et. al. (2001) also use the ENADID survey to show that Mexican migrants have become more negatively selected over time after controlling for cohort effects (though still predominantly drawn from the middle of the education distribution).

self selection, but ultimately, these studies are subject to legitimate criticism related to problems of measurement and data availability. In this paper, we add to this literature by studying the question of immigrant selection using arguably the most appropriate data available.[6] As noted earlier, the MxFLS is not subject to some of the potentially serious problems that have affected previous work. Specifically, the MxFLS allows us to identify all migrants, even those that have permanently moved to the U.S. Studies that use the Mexican Census (Ibarraran and Lubotsky 2007), the MMP (Orrenius and Zavodny 2005), the ENADID (McKenzie and Rapoport 2007), or the ENET (Fernández-Huertas Moraga, 2008) do not observe most permanent migrants. The MxFLS also avoids the potential undercount of Mexican migrants in the U.S. Census, which is the data source used by Chiquiar and Hanson (2005). Further, the MxFLS does not depend on recall to identify migrants and time of migration like the MMP. Perhaps most importantly, the MxFLS has information on earnings of migrants before they have left Mexico, as well as proxies for the costs of migration. Thus, we can examine selection on both observed and unobserved skills of Mexican migrants while attempting to control for the costs of migration, something which has not been done by previous studies.[7]

3 Theory

In order to motivate the empirical analysis, this section presents a model of migration based on Borjas (1987, 1991).[8] In our version of the model, mobility costs are allowed to vary with schooling and with unobserved skills.[9] Individuals from Mexico, indexed by 0, choose whether or not to

[6] Rubalcava, et al. (2008) use these same data to examine the health selectivity of migrants to the US. They also report the absence of any differential selection on education when measured as a linear function of years of schooling. This is consistent with our findings since the pattern of intermediate selection is only apparent when allowing for a non-linear functions of schooling.

[7] Gould and Moav (2008) examine selection on both observed and unobserved skills for Israeli migrants. They find positive selection on education and intermediate selection on residual wages. However, it is important to note that the relative returns to education and unobserved skills are very different between the United States and Israel, as compared to between Mexico and the United States.

[8] Grogger and Hanson (2008) provide an alternative model of immigration where the decision to migrate is based on a linear (utility) function of wages rather than log (utility) function of wages. The key difference between their model and the model presented in this paper is that migration decisions depend on differences in wage levels between source and destination country in addition to differences in the return to skill in the source and destination country.

[9] See McKenzie and Rapoport (2007) for a similar model that allows migration costs to be correlated with skill.

migrate to the United States, indexed by 1, on the basis of their potential earnings net of moving costs. For simplicity, we treat this as a one-time decision. Residents of Mexico earn a wage given by:

$$\ln(w_0) = \mu_0 + \delta_0 s + \varepsilon_0, \tag{1}$$

where w_0 is the wage in Mexico, μ_0 is the base wage in the absence of any schooling, s is the level of schooling, δ_0 is the return to schooling in Mexico, and ε_0 is independent of s and normally distributed with mean zero and variance σ_0^2. If the population of Mexico were to migrate to the United States, they would earn a wage given by:

$$\ln(w_1) = \mu_1 + \delta_1 s + \varepsilon_1, \tag{2}$$

where w_1 is the wage of Mexican migrants in the United States, μ_1 is the base wage without schooling, δ_1 is the return to schooling for Mexicans in the United States, and ε_1 is independent of s and normally distributed with mean zero and variance σ_1^2. The random variables ε_0 and ε_1 may be correlated with correlation coefficient, ρ, which is the degree to which unmeasured skills of Mexican migrants are similarly valued in both Mexico and the United States. The education distribution of the population of Mexico can be written as:

$$s = \mu_s + \varepsilon_s \tag{3}$$

where ε_s is normally distributed with mean zero and variance σ_s^2.

Equations (1) and (2) completely describe the earnings opportunities facing a potential migrant from Mexico. Combining these two equations and assuming the presence of migrations costs, denoted in time-equivalent units by π, implies that a resident from Mexico will migrate to the United States if

$$I = \ln(w_1) - \ln(w_0) - \pi = (\mu_1 - \mu_0) + [(\delta_1 - \delta_0)\, s - \pi] + (\varepsilon_1 - \varepsilon_0) > 0. \tag{4}$$

We depart from the assumption of constant migration costs and suppose that mobility costs can be written as:

$$\pi = \mu_\pi + \delta_\pi s + \varepsilon_\pi, \tag{5}$$

where μ_π is the mean level of mobility costs in the population, δ_π captures the relationship between schooling and mobility costs, and ε_π is normally distributed with mean zero and variance σ_π^2. The random variable ε_π may be correlated with ε_1 and ε_0, with the correlation coefficients given by $\rho_{\pi 0}$ and $\rho_{\pi 1}$ respectively. This implies that the rate of emigration from Mexico to the United States is given by:

$$\begin{aligned} P &= \Pr\left\{(\varepsilon_1 - \varepsilon_0) + (\delta_1 - \delta_0 - \delta_\pi)\,\varepsilon_s + \varepsilon_\pi > -(\mu_1 - \mu_0) + (\delta_1 - \delta_0 - \delta_\pi)\,\mu_s - \mu_\pi\right\} \\ &= 1 - \Phi(z^*) \end{aligned} \tag{6}$$

where $t = (\varepsilon_1 - \varepsilon_0 + \varepsilon_\pi) + (\delta_1 - \delta_0 - \delta_\pi)\,\varepsilon_s$, and $z^* = -\left[(\mu_1 - \mu_0) + (\delta_1 - \delta_0 - \delta_\pi)\,\mu_s - \mu_\pi\right]/\sigma_t$. Given this framework, it is straightforward to show that the conditional expectation of schooling for Mexicans who migrate to the United States is:

$$E(s \mid I > 0) = \mu_s + \frac{\sigma_s^2}{\sigma_t^2}(\delta_1 - \delta_0 - \delta_\pi)\,\lambda, \tag{7}$$

where $\lambda = \phi(z^*)/P$. In other words, the schooling level of Mexican migrants will be equal to the mean level of schooling of non-migrants plus the selection effect, which depends on the difference in the return to schooling between the U.S. and Mexico, *and* on the relationship between migration costs and schooling. If we assume that migration costs are constant across different levels of schooling, so that $\delta_\pi = 0$, the standard comparative statics would apply: there would be positive selection in schooling when $(\delta_1 - \delta_0) > 0$ because highly educated individuals from Mexico would be more highly rewarded in the U.S. labor market, and negative selection in schooling when $(\delta_1 - \delta_0) < 0$ because these highly educated individuals would earn more by remaining in Mexico.

However, if migration costs are inversely related to schooling, such that $\delta_\pi < 0$, we may observe positive selection even in the case that $(\delta_1 - \delta_0) < 0$. Note that by conditioning on migration costs, $E(s \mid \pi, I > 0)$, we return to the case where the degree of selection depends only on the relative returns to schooling between Mexico and the United States.

Now consider selection based on unobserved skill, ε. We can write the expectation of log wages in Mexico conditional on the observed level of schooling and the decision to migrate as follows:

$$E(\ln w_0 \mid s, I > 0) = \delta_0 s + \frac{\sigma_1 \sigma_0}{\sigma_t} \left[\left(\rho - \frac{\sigma_0}{\sigma_1} \right) - \rho_{\pi 0} \frac{\sigma \pi}{\sigma_1} \right] \lambda. \tag{8}$$

Hence, the nature of selection on unobserved skill depends on whether the dispersion in the return to unobserved skill in the United States is greater than the dispersion in the return to unobserved skill in Mexico, σ_0/σ_1, the correlation between the returns to unobserved skill in the United States and Mexico, ρ, as well as the correlation between migration costs and unobserved skill in Mexico, $\rho_{\pi 0}$. Again, if we were to assume that migration costs are uncorrelated with unobserved skill, so that $\delta_\pi = 0$, the standard comparative statics would apply. For a sufficiently high value of ρ, we should observe positive selection in unobserved skill when there is greater inequality of the returns to unobserved skill in the United States than in Mexico, and negative selection when there is greater inequality of the return to unobserved skill in Mexico than in the United States.[10]

However, if migration costs are correlated with unobserved skill, then the type of selection generated may be in either direction. A positive correlation between migration costs and unobserved skills increases the likelihood of negative selection whereas a negative correlation between migration costs and unobserved skill increases the likelihood of positive selection. Note, once more, that by conditioning on migration costs, $E(\ln w_0 \mid \pi, s, I > 0)$, we return to the case where the degree of selection depends only on whether the dispersion in the return to unobserved skill in the United States is greater than the dispersion in the return to unobserved skill in Mexico, σ_0/σ_1, and the correlation between the returns to skills in the United States and Mexico, ρ.

[10] Note that a small or negative value for ρ implies that migrants will have below-average earnings in Mexico but above-average earnings in the United States (what Borjas refers to as "refugee selection").

To summarize, this simple model based on Borjas (1987) has a few, well-known empirical implications. First, the nature of the selection of Mexican migrants on observed skills depends on the difference in the return to skill (e.g., education) in the U.S. and Mexico, and the correlation between the costs of migration and observed skill. Consider the case in which skill is measured by education. Previous studies, and our own estimates, from the Mexican and U.S. Censuses have documented that the return to education is higher in Mexico than in the U.S. Thus, holding costs of migration constant, theory predicts that Mexican migrants will be negatively selected on education. However, if costs of migration are not held constant, then Mexican migrants may be negatively or positively selected on education (or selected from the middle of the education distribution). Similar arguments apply to more multi-dimensioned measures of skill such as those that also include age and, in our case, cognitive ability. The other prediction from the model is that, holding costs of migration constant, selection on unobserved skill depends on the relative dispersion of the returns to unobserved skills in the U.S. and Mexico. Again, allowing costs of migration to differ by skill makes the prediction ambiguous. The following section describes the data and the empirical strategy used to test this model.

4 Data and Empirical Strategy

4.1 Data

The Mexican Family Life Survey is a longitudinal household survey that is representative at the national, urban-rural and regional level. The first round of the survey was conducted from April to July 2002 and collected information from a sample of approximately 8,400 households or 35,000 individuals across 150 different communities in 16 of Mexico's 32 states.[11] The baseline survey included both household information and individual-level information for all members living in the household. The survey covered many areas: educational attainment, labor market participation and earnings, past migration experiences, assets, credit availability, and cognitive ability. The second

[11] See http://www.ennvih-mxfls.org/ for more details.

round of the survey was begun in mid-2005 and completed in 2006, with a 90 percent re-contact rate. Those who had migrated to the United States were contacted at a rate of over 91 percent. Although information from the second round interviews of migrants to the United States has not yet been released, an indicator for whether the individual has moved to the United States is available in the public-use data; note that all migrants should be identified even if they could not be found in the second round.[12]

We restrict the sample to males aged 21 to 65 in the first round of the MxFLS.[13] Table 1 presents unweighted summary statistics for this sample. Panel A indicates that the fraction who migrated to the United States over the three year period prior to the second round of the MxFLS is 3.63 percent, or 295 out of 8,116 men.[14] This rate of migration is consistent with other sources. Hanson and McIntosh (2010) used consecutive Mexican Censuses to estimate migration rates and reported 10-year migration rates of approximately 12-14 percent for young men in their 20s and 5-10 percent for men in their 30s. In our sample, the 3-year migration rates for men in their 20s and 30s are 5.6 percent and 3.3 percent respectively, and are therefore consistent with those reported by Hanson and McIntosh (2010). Passel and Cohn (2009) used CPS data to estimate that approximately 550,000 Mexican immigrants arrived in the US each year between 2003 and 2006. This figure represents about 1.1 percent of the non-elderly adult population of Mexico. In the MxFLS, approximately 2.8 percent of the baseline adult (male and female) sample migrated between 2002 and 2005, which is roughly equivalent to the annual rate reported by Passel and Cohn.

The variables in Panels B and C are from the first round of the MxFLS in 2002, reported by the head of household (or other knowledge person in the household) and the individual, respectively.

[12] In cases where entire households were no longer present at the original address, the interviewers searched for the final destination of that particular household. This was aided by information collected in the first round of the survey where in the case of the household were to move, the address and telephone number of close relatives (including those living in the United States) could assist in providing current household information.

[13] An analysis of female migration is complicated by the fact that the migration decision is more likely to be tied to spouse and labor market participation is substantially lower making it more difficult to compare wages. This latter point also applies to the analysis of men younger than 21 years of age (as shown in Appendix B).

[14] This includes 217 men who were residing in the United States and 78 men who moved to the United States but returned to Mexico by time of the second round. We do not exclude return migrants because they migrated subsequent to the baseline survey and their decision to return to Mexico is likely to be based partly on ex-post realizations of their experiences in the United States.

The main variables from Panel B used in the empirical analysis are age, marital status, and the total value of household assets. But we also show that average years of schooling and the proportion working in the sample, as reported by the head of household, are very similar to reports by the individuals themselves. The proportion of our sample living in rural areas is about 50 percent which is somewhat higher than the 41 percent of an analogous sample in the 2000 Mexican Census. However, the average years of schooling in our sample is almost identical to the 7.27 years of schooling reported for a similar sample of men aged 21 to 65 in 2000 Mexican Census. The MxFLS also provides information on cognitive ability, a dimension of skill that is usually unmeasured in most data sets. Cognitive ability is measured as the number of correct responses to 12 multiple-choice questions of a Raven's (1983) Progressive Matrices test. In addition, the MxFLS contains unique information on plausible proxies for the costs of migration (presence of relatives in the US, past migration experience, US documentation) and credit constraints (household assets, individual savings, and availability of credit).

We focus on individual reports of annual earnings because these include several important categories, such as profit-sharing and Christmas bonuses, which are omitted in individual reports of monthly earnings.[15] Moreover, since all of the individual interviews were conducted between April and July, monthly earnings may be subject to seasonal variation. We include earnings in the main job as well as any earnings received from a secondary job, and use proxy reports to complete any missing information.[16] Nevertheless, we will assess the sensitivity of our findings to excluding proxy reports, excluding imputed values, and calculating earnings and wages on the basis of monthly earnings. We also calculate hourly wages based on the total number of hours worked throughout the year (as number of weeks worked during the year multiplied by the usual hours worked per week).[17] The average hourly wage based on annual earnings for our sample is about

[15] Federal law requires firms to participate in a profit sharing program in which employees receive 10 percent of the firm's annual profits. Firms are also required to pay a year-end Christmas bonus (Aguinaldo) to all employees equivalent to at least two-weeks pay.

[16] Proxy reports account for approximately 10-15 percent of observations for most variables. For individuals who are working but do not report positive earnings, we impute earnings based on age, years schooling, marital status, urban status, state fixed effects and earnings information as reported by the head of household.

[17] We drop outliers that appear to be data errors. Specifically, we trim individuals with hourly wages less than 0.1 and greater than 1000 pesos. This accounts for approximately 0.5 percent of the sample.

17 pesos which is similar to the 18 pesos an hour based on monthly earnings in the 2000 Mexican Census. Appendix A describes the construction of all the variables in Table 1 in greater detail.

4.2 Empirical Strategy

The MxFLS data is ideally suited to test the predictions of the Borjas (1987, 1991) model in the context of Mexican migration to the United States. To do so, we first describe the nature of selection of migrants separately for three observed measures of skills: education, age and cognitive ability. In particular, we examine the proportion of migrants from different parts of the education distribution, age distribution, and cognitive ability distribution. We also compare the wage distributions of migrants and non-migrants. This exercise essentially replicates earlier studies examining the observable characteristics of Mexican migrants to the U.S., but takes advantage of a representative survey that does not undercount certain types of Mexican migrants.

The preceding analysis considers each observable skill separately and ignores the price of these skills in the labor market. We therefore proceed to price out observed skills and analyze the nature of selection using an index of observed skills, as measured by predicted wages. Predicted wages are obtained using various combinations of observed skills, beginning with just education, adding age, and finally introducing a measure of cognitive ability. The wage models are estimated using a saturated model that include all interactions between the different aspects of observed skill. The difference between migrants and non-migrants in observed skill are assessed using the following regression model:

$$\Pr\left(m_i = 1\right) = \alpha_0 + \alpha_1 O_i + \epsilon_i, \tag{9}$$

where m_i is an indicator variable for having migrated to the United States and O_i is an index of observed skills. We carry out this analysis by collapsing our measures of education, age, and cognitive ability into different categories. For example, we construct the index of observable skill based on education, age, and cognitive ability by regressing wages on the full set of interactions between j education, k age, and l ability categorical variables:

$$\ln w_i = \sum^{j} \gamma_j educ_{ji} * \sum^{k} \lambda_k age_{ki} * \sum^{l} \theta_l ability_{li} + \eta_i, \tag{10}$$

and forming predicted wages on the basis of the estimated coefficients so that $O_i = \Sigma\widehat{\gamma}_j educ_{ji} * \Sigma\widehat{\lambda}_k age_{ki} * \Sigma\widehat{\theta}_l ability_{li}$. If there is negative selection of migrants in terms of observable skills, the coefficient on our index of observable skills would be negative indicating that migrants are over-represented among Mexicans with below-average skills and underrepresented among the Mexicans with above-average skills. In fact, we consider flexible specifications of O_i by including indicator variables for different quantiles of the predicted wage distribution that allow us to pick up non-linear patterns of selection.

We then conduct a similar analysis of the selection of migrants by unobserved skill by estimating an analogous regression model:

$$\Pr\left(m_i = 1\right) = \beta_0 + \beta_1 U_i + \zeta_i, \tag{11}$$

where $U_i = \widehat{\eta}_i$ is unobserved skill measured as the residual of wages when regressed on our set of observable skills. Again, we consider flexible specifications of U_i by including indicator variables for different quintiles of the residual wage distribution that allow us to pick up non-linear patterns of selection.[18] Note that it is only because the MxFLS has earnings information for migrants while they are still in Mexico that we can conduct the analysis of selection on unobserved skill.

Finally, we re-assess the nature of selection on observed and unobserved skills with additional controls for the costs of migration, the availability of credit, as well as community fixed effects. According to the theory, we expect to see greater evidence of negative selection when migration costs are included than when they are not included. To the extent that some individuals face binding liquidity constraints, adding measures of credit availability may also affect the nature of selection. Community fixed effects should capture any differences across communities that may be correlated with migration decisions. While certain communities may be associated with different costs of

[18] For simplicity, we estimate equations 9 and 11 using a linear probability model but our results are robust to probit and logit specifications.

migration, there are likely to be important differences in terms of relative returns to observed and unobserved skills across communities (as documented by Ibarraran and Lubotsky, 2007).

5 Results

5.1 Selection of Migrants on Age, Education, Ability and Earnings

We begin by describing the nature of selection of migrants by education, age, cognitive ability and earnings. Table 2 shows the proportion of migrants and non-migrants across five education categories, five age categories, four categories of cognitive ability, as well as earnings and wage quintiles. On average, migrants are younger than non-migrants, and there is a steep gradient in the likelihood of migrating by age. Compared to non-migrants, migrants are significantly over represented among those aged 21 to 29, and significantly under represented among those aged 48 to 65. For example, 44 percent of all migrants are aged 21 to 29 as compared to 28 percent of non-migrants.

With respect to education, Table 2 indicates that migrants are significantly over represented in the middle of the education distribution, similar to the findings in Chiquiar and Hanson (2005) and Orrenius and Zavodny (2005). Among migrants, 32 percent have seven to nine years of schooling compared to 25 percent of non-migrants. Panel A of Figure 1 shows the probability of migrating disaggregated by each year of schooling, with the solid line depicting a quadratic fit to the probability of migrating (estimated at the individual level). The general pattern of results is the same—greater probability of migrating among those in the middle of the education distribution than those at either the top or bottom of the distribution—but there is somewhat less uniformity to the pattern than suggested by estimates in Table 2.[19]

In contrast to age and education, migrants differ very little from non-migrants on cognitive ability. Table 2 reveals that the proportions of migrant and non-migrants are virtually identical across different categories of cognitive ability. Panel B of Figure 1 illustrates the probability of

[19] Specifically, among those with zero years of schooling, the migration probability is approximately four percent and about equal to the probability of migration among those with six, eight and nine years of schooling, and greater than the probability of migrating among those with ten, twelve and more years of schooling.

migrating for a more disaggregated measure of cognitive test score (and again, the solid line depicts the quadratic fit to the proportion of migrants by cognitive ability score). The absence of any difference between migrants and non-migrants with respect to cognitive test scores is surprising given the findings for education and the relatively high positive correlation between years of schooling and cognitive test score (r=0.49).

Migrants are most heavily represented in the second quintile of the earnings distribution regardless of whether earnings are measured on an annual or hourly basis. According to Table 2, approximately 25 percent of migrants are in the second quintile compared to 19 percent of non-migrants. Migrants are also significantly less likely than non-migrants to be found in the top quintile of annual earnings. Panels A and B of Figure 2 show the distribution of (log) earnings for migrants and non-migrants on an annual and hourly basis respectively.[20] Both panels show that migrants are more likely to be found in the middle of the earnings distributions than non-migrants. Panels C and D plot the difference in the migrant and non-migrant distributions for (log) earnings and (log) hourly wages, respectively. This finding is similar to that of education, and is consistent with the strong influence of education on earnings, even though wages include other components of skill. Note that there are no significant differences in the probability of working between migrants and non-migrants.

To summarize, estimates in Table 2, and the relationships presented in Figures 1 and 2, show that Mexican migrants differ from non-migrants along observable dimensions of skill, and that migrants tend to be overrepresented in the middle of the education and earnings distributions. Surprisingly, migrants and non-migrants tend to have similar distributions of cognitive test scores. Migrants are also over represented among those 21 to 29 years of age and under represented among those aged 48 to 65. The finding of selection from the middle of the education and earnings distributions is similar to some previous research, and is somewhat of a puzzle given that the returns to education in Mexico are generally higher than in the U.S.[21] Although the pattern of

[20] These densities are estimated using an Epanechnikov kernel and Silverman's (1986) optimal bandwidth, $h = 0.9\widehat{\sigma}N^{-1/5}$ with $\widehat{\sigma} = \min\left\{S, \frac{IQR}{1.349}\right\}$ where S is the standard deviation, IQR is the interquartile range, and N is the number of observations.

[21] Our calculations from the 2000 U.S. and Mexican Censuses indicate that returns to education in Mexico are

selection by age is also potentially inconsistent with a simple model of migration, given that the returns to age in Mexico are relatively higher among the middle-aged than among younger men as compared to the U.S., the longer time horizon and potentially lower costs (in terms of family attachments, etc.) probably explains much of why younger workers are so much more likely to migrate. Table 2 also presents differences between migrants and non-migrants for a number of variables that can serve as proxies for the costs of migration: marital status, whether a person has a family member living in the U.S., whether a person has previously visited the U.S. whether a person has U.S. documentation (visa, greencard, or passport), and whether a person has considered moving to the U.S. All these variables are significantly different between migrants and non-migrants suggesting that these costs are important determinants of migration.

5.2 Selection of Migrants on an Index of Observed Skills

The preceding analysis considered each observable skill separately and ignored the price of these skills in the labor market. In this section, we price out observed skills and examine whether migrants and non-migrants differ on an index of observed skills, measured by predicted wages. As explained in the section 4.2, we construct this index of observed skill by regressing observed wages on various combinations of age, education and cognitive test scores and forming a predicted measure of wages (where age can be viewed as a proxy for experience). We then assess whether the probability to migrate differs by the observed skill index, as measured by quintiles of predicted wages (i.e., observed skills).[22] We use two indices of predicted skill – one based on predicted annual earnings and one based on predicted hourly wages – and construct each index three ways: using only education to predict wages, using age and education (including all interactions) to predict wages, and using age, education and cognitive test scores (including all interactions) to predict wages.

higher than in the U.S. for the most educated and lower than in the U.S. for the least educated (0-3 years). Indeed, the differences in the return to education between Mexico and the U.S. are increasing (in favor of Mexico) in years of schooling.

[22] We have examined alternative specifications, such as quartiles and tertiles, with similar results. Our findings also hold with parametric specifications that include non-linear (i.e. quadratic) terms in predicted wages.

Table 3 presents the regression estimates of the association between the probability of migrating and the index of observed skill distribution, as measured by predicted wages. Consider the estimates in column (1) where we use only education to predict wages. For both annual earnings and hourly wages, the estimates indicate that migrants are significantly more likely to come from the 3rd quintile of the observed skill distribution than the 1st (bottom) quintile, and significantly less likely to come from the 5th quintile than the 1st quintile of the observed skill distribution. The magnitudes of the differences are relatively large. The probability of migrating is 0.013 percentage points, or 37 percent of the mean, greater in the 3rd quintile than the 1st quintile. These estimates are consistent with estimates in Table 2 showing that migrants were significantly over represented among those with seven to nine years of schooling and significantly under represented among those with 12 or more years of schooling. This implies that restricting attention to the sample of wage earners does not change the main finding of selection from the middle of the education distribution. Further note that we get similar results with the somewhat smaller sample of men with wage observations as with the larger sample of men with information on years of schooling. This confirms that labor market participation differences are not driving the findings for predicted wages.

Columns (2) and (3) show estimates of the association between the probability of migrating and observed skill when the observed skill indices are constructed by using age and cognitive test scores in addition to education to predict wages. In each case, migrants are significantly more likely to come from the 3rd and 4th quintiles than the 1st quintile of the observed skill distribution. The magnitudes of the differences here are also relatively large. The probability of migration is 0.015 (44 percent) to 0.023 (67 percent) greater in the 3rd and 4th quintiles than the 1st quintile. Overall, these estimates, based on more comprehensive measures of observed skill, continue to show that migrants are drawn disproportionately from the middle of the observed skill distribution.

5.3 Selection of Migrants on an Index of Unobserved Skills

One of the advantages of the MxFLS is that we observe earnings in Mexico for both migrants and non-migrants, and we can use these earnings data to assess the nature of selection of migrants on

unobserved skill as well as observed skills, which has not been previously studied. The unobserved skill index is based on residual earnings – the portion of earnings orthogonal to measured aspects of skill such as age, education and cognitive test scores. Thus, we can obtain estimates of the association between the probability to migrate and the distribution of unobserved skills, as measured by quintiles of residual earnings. Table 4 presents these estimates and it is structured analogously to Table 3.

In contrast to our findings for observed skill, estimates in Table 4 indicate that the probability of migrating is not significantly related to unobserved skill. These results imply that, for the most part, migrants and non-migrants have similar distributions of unobserved skill. The exception is among those in the 5th (top) quintile of unobserved skill, but only when unobserved skill is based on residual annual earnings. In this case, the probability of migration among those in the 5th quintile is significantly lower, approximately 50 percent of the mean, than in the 1st quintile. If we consider the results in Tables 3 and 4 simultaneously, we can infer the source of the selection from the middle of the wage distribution revealed in Figure 2. The over representation of migrants in the middle of the wage distribution is due to the over representation of migrants in the middle of the observed skill distribution, as there is relatively little difference in the distribution of unobserved skill between migrants and non-migrants.

5.4 Proxies for the Costs of Migration and Differential Returns to Skill

The pattern of selection of migrants from the middle of the education distribution is inconsistent with a simple model of migration that assumes constant costs of migration. The higher observed return to education in Mexico compared to the U.S. implies negative selection, so we expect migrants to be over represented in the bottom of the education distribution and under represented at the top of the education distribution. However, it is unlikely that the costs of migration are constant, and there are plausible reasons to expect the costs of migration to be negatively related with education (Chiswick, 1999). For example, more educated individuals may find it easier to navigate the bureaucratic procedures necessary to migrate legally to the U.S. or cover the fixed costs necessary

to enter the U.S. (either legally or illegally). If the costs of migration are negatively related to education, then ignoring such costs will make it more likely to observe positive (or less negative) selection of migrants by education. The pattern of selection across the observed skill distribution will likewise be affected by differential costs. Moreover, the Mexico-U.S. difference in the return to observed skill may not be constant throughout the skill distribution. If the return to skill varies over the distribution, then we would also expect the nature of the selection of migrants to vary over the skill distribution.[23]

In this section, we examine the skill distributions of migrants and non-migrants conditional on proxy variables for the cost of migration, and conditional on differences in the return to observed skill between Mexico and the U.S. Table 5 shows the probability of migrating at different quintiles of the observed skill distribution with these additional controls for the full index of observed skill based on education, age, and cognitive ability.[24] Panel A presents estimates for observed skill constructed using annual earnings, and Panel B presents estimates for observed skill constructed using hourly wages. Estimates in column (1) of each panel are taken directly from Table 3 and are presented to facilitate comparison with other estimates in Table 5.

Estimates in column (2) of Table 5 are based on regressions that include proxy variables for the costs of migration. These variables include marital status, whether a person has a family member living in the U.S., whether a person has previously visited the U.S. whether a person has U.S. documentation (visa, greencard, or passport), and whether a person has considered moving to the U.S. In both panels, the addition of controls for the costs of migration substantially changes the pattern of selection on observed skills. Clearly, these proxy variables for costs of migration are systematically related to the observed skill distribution and probability of migration, and the finding of selection from the middle of the observed skill distribution is not robust to the inclusion of such controls. Indeed, all of our proxies for the costs of migration are highly significant in the expected directions (not shown). Comparing estimates in column (2) to those in column (1)

[23] Note, however, that if the return to observed skill is monotonically increasing in skill, then the selection from the middle of the skill distribution reported in Table 3 cannot be solely explained by variation in the Mexico-U.S. differences in the return to skill.

[24] The results in Table 3 are similar for an index of observed skill based on just education, or on education and age.

reveals that the addition of proxy variables for the costs of migration significantly decreases the association between the probability of migration and being in the 3rd and 4th quintiles of the observed skill distribution (versus the 1st quintile). In fact, none of the estimates in column (2) are statistically significant and most of the estimates are relatively small when measured against the mean, although somewhat imprecise. Thus, including controls for the costs of migration provides far less evidence for the selection of migrants from the middle of the observed skill distribution, and raises the possibility that more complete controls for the costs of migration would actually reverse the pattern of selection.

In column (3), we add controls for household assets and access to credit (whether individuals know of any person or place where one can borrow or ask for a credit, and whether they have any savings). Some have argued that the inability to finance migration is an important determinant of Mexican migration, and failure to consider this may affect the pattern of selection (Orrenius and Zavodny 2005). Estimates in column (3) in Panels A and B of Table 5 are very similar to those in column (1). Thus, for our measures of assets and credit, there is little evidence that these variables play an important role in the nature of selection of migrants.

The next specification in Table 5 adds community fixed effects. The main motivation for this specification is evidence regarding community-specific differences in the return to skill (Ibarraran and Lubotsky 2007). In addition, communities may differ in the extent of migration networks and other characteristics that are likely to alter the costs of migration (McKenzie and Rapoport 2007). It is unclear which of these two possibilities may be more important. Column (4) presents estimates for this specification indicating that the probability of migration is significantly higher for those in the 2nd to 5th quintiles than those in the 1st quintile. The magnitudes of the estimates continue to suggest that migrants are disproportionately drawn from the middle of the observed skill distribution, although the pattern of the associations between the probability of migration and the observed skill distribution appears to be more of a step function than the inverted U-shape pattern seen in column (1). However, it is clear that observed skill is highly correlated with community fixed effects.

Theory predicts that the probability of migration depends on differences in the return to observed skill between Mexico and the U.S. Accordingly, we calculated the return to observed skill based on age and education for Mexican migrants in the U.S. (using the 2000 U.S. Census) and residents of Mexico (using the 2000 Mexican Census). We include a full set of interactions between our five education and five age categories. We then merge the differential in the estimated return to skill by age and education categories, and include this differential in the regression model (while clustering on age*education). The resulting estimates are shown in column (5) of Table 5. The coefficient on the Mexico–U.S. differential in the return to skill is positive and significant in both panels, which is consistent with theory. A one standard deviation (0.44) increase in the relative return to skill in Mexico versus the U.S. is associated with a 0.023 percentage point (69 percent of mean) increase in the probability of migration. Moreover, estimates in column (5) are quite similar to those in column (4); the probability of migration is significantly higher for those in the 2nd to 5th quintiles of the observed skill distribution than those in the 1st quintile. The similarity of estimates in columns (4) and (5) suggest that community-specific fixed effects are be picking up differences in the return to observed skill, consistent with the findings in Ibarraran and Lubotsky (2007).

Finally, the estimates in column (5) suggest that differences in the costs of migration are strongly correlated with observed skill and the probability of migration. This inference is plausible given that the differential in the Mexico-U.S. returns to skill are included in the model. Thus, the selection on observed skill revealed by estimates in column (5) is likely being driven by differences in the costs of migration. Indeed, this hypothesis is supported by estimates in column (6) that include both the differential in the Mexico-U.S. return to skills and proxy variables for the costs of migration. Estimates in column (6) are substantially less positive than estimates in column (5), and the differences between estimates in column (6) versus those in column (5) are similar to differences between estimates in columns (2) versus column (1). In both cases, the addition of proxy variables for the costs of migration weakens the pattern of selection from the middle of the observed skill distribution, and provides more evidence in support of the standard prediction from theory.

We also conduct a similar analysis as that represented by Table 5, but using residual wages as an index of unobserved skill. Estimates from this analysis are presented in Table 6. In general, the estimates in Table 6 are small (relative to the mean) and statistically insignificant. There is little evidence of selection on unobserved skill, which is similar to the conclusions we drew from Table 4. The addition of proxy variables for the costs of migration and variables measuring assets and access to credit had little effect on estimates of the associations between the probability of migrating and quintiles of the unobserved skill distribution. The only significant associations are those between belonging to the 5th (top) quintile of the distribution of unobserved skill based and the probability of migration. Those in the top quintile of the unobserved skill distribution are significantly less likely to migrate than those in the bottom quintile, although this is only true in the top panel when unobserved skill is measured by residual annual earnings. The addition of community fixed effects eliminates the significant association between being in the top quintile of the unobserved skill distribution and the probability of migration.

In Appendix Table 1, we explore whether our results are robust to different specifications of earnings and wages in the case of observed skill. Specifically, we assess the sensitivity of our findings to excluding information from proxy reports (by a knowledgeable member of the household), to excluding imputed values for earnings and wages for individuals who are working but do not report positive earnings, and to calculating earnings and wages on the basis of monthly earnings. Our results are generally robust to these alternative specifications. We observe selection from the middle of the observed skill distribution when excluding controls for the costs of migration (columns 1, 3, and 5), and more negative selection once controls for the costs of migration are included (columns 2, 4, and 6). Although not shown here, the results in Table 6 are also robust to these alternative specifications of earnings and wages.

6 Discussion

How do our findings relate to those reported in previous studies? In models that do not include variables measuring the costs of migration, we find evidence for "intermediate" selection of Mexican

migrants on education and on an index of observed skill. These findings are similar to results reported by Chiquiar and Hanson (2005), Orrenius and Zavodny (2005), and, with respect to education, Fernández-Huertas Moraga (2008). However, Ibarraran and Lubotsky (2007) found negative selection of migrants on education, and Fernández-Huertas Moraga (2008) found negative selection of urban migrants on earnings (but positive selection of rural migrants on earnings). Ibarraran and Lubotsky (2007) argue that their finding is different from Chiquiar and Hanson (2005) because it corrects for certain data problems, specifically, the undercount of migrants in U.S. Census and over reporting of education of migrants in U.S. Census. Fernández-Huertas Moraga (2008) also argues that the difference between his findings and the findings of Chiquiar and Hanson (2005), at least in the case of urban migrants, is because of the undercount of migrants in the U.S. Census.[25] However, our data are not subject to this undercounting problem or to the problem of over reporting of education, and yet we find "intermediate" selection on both education and an index of observed skill.

An explanation for the difference between our findings and those of Fernández-Huertas Moraga (2008) may be related to differences in the samples used in each of our studies. Both the MxFLS and the ENET are nationally representative samples, although the ENET misses migrants in households in which all members have migrated to the United States. According to our data, approximately 13 percent of migrants moved to the United States together with their entire households and they appear to be positively selected on education and wages. Nevertheless, because of the relatively small fraction of migrants affected, this is unlikely to account for the large differences in our results. More important is the inclusion of migrants aged 16 to 20 in the analysis conducted by Fernández-Huertas Moraga (2008). We exclude these younger migrants in our sample because they have very low rates of labor market participation. The fraction of young men aged 15 to 20 who report that they are currently working is only 54 percent. There are also significant differences in labor market participation by migrant status: those men who ended up migrating were 11 percentage points (or

[25] In addition, Fernández-Huertas Moraga (2008) notes that part of the discrepancy between his results and those of Chiquiar and Hanson (2005) may be due to negative selection on unobserved skill, but only offers indirect evidence in support of this hypothesis.

22 percent) more likely to work prior to migrating as compared to their counterparts who remain in Mexico. Consequently, there is more evidence of negative selection on earnings and wages for these younger men than in our main sample of 21 to 65 year olds. However, there are no substantial differences between these samples in the pattern of selection on education. This suggests that including younger men in the wage analysis may lead to substantial bias due to non-participation and might explain the difference between our findings and those of Fernández-Huertas Moraga (2008).

Our findings are also related to the debate over the role of the costs of migration that has been prominent in the previous literature. For example, McKenzie and Rapoport (2007) argue that their finding of negative selection on education in communities with large migrant networks and positive selection on education in communities with smaller migrant networks illustrates the important role of costs of migration, and may be an explanation of inconsistent findings in other studies. However, this interpretation depends on the relationship between the costs of migration due to education and those due to migrant networks. Theory suggests that stratifying the sample by proxy variables for migration costs such as the size of the migrant network should yield greater evidence for negative selection on observed skill in both types of communities if these costs are unrelated to the costs that can be overcome with education. Only if the presence of migrant networks eliminates all or most costs, and education can compensate for the absence of migrant networks, would we expect to find negative selection in one type of community and positive selection in the other. Similarly, Fernández-Huertas Moraga (2008) reported negative selection on skill of migrants in urban areas, but positive selection on skill of migrants in rural areas. An explanation for this pattern that depends just on the costs of migration must assume that urban locations eliminate all or most costs, and that education can compensate for such costs in a rural setting. Our results provide greater clarity on this issue. We showed directly that costs of migration play an important role in the migration decision, and that partially controlling for these costs results in significantly less evidence of "intermediate" selection on skills. Sample sizes prevent us from stratifying the sample based on urban/rural status, but as we showed with respect to community fixed effects, urban/rural status

and the size of migration network may be correlated with both differential returns to migration and different costs of migration. So stratifying the sample on the basis of community is not necessarily a good test of the role of migration costs. Here, we used proxy variables for the costs of migration directly, and found evidence consistent with theory.

7 Conclusion

In this paper, we presented new evidence of the nature of selection of male, Mexican migrants to the U.S. Our analysis was based on a new data source, the Mexican Family Life Survey (MxFLS), which is ideally suited to study the selection of Mexican migrants. Most importantly, the MxFLS has information on the education, cognitive ability and earnings prior to migration of Mexican migrants, including those in which the entire household migrated to the U.S. These data allow us to examine the nature of selection on a more comprehensive measure of observed skill than that used in previous studies, and to analyze the nature of selection on unobserved skill, which is something no prior study has done directly. Furthermore, the MxFLS contains variables that are good proxies for the costs of migration, and we use this information to help identify the pattern of selection of migrants on skill.

In analyses that ignore the cost of migration, we found that male, Mexican migrants were more likely to come from the middle of the distribution of observed skills regardless of whether observed skill was measured by education; age and education; or age, education and cognitive ability. This finding was robust to whether we measured observed skill using an index of skill based on annual earnings or hourly wages. In contrast, we found little consistent evidence of selection of migrants on unobserved skill as measured by residual earnings and wages.

Our findings are quite different once we partly control for proxy variables for the costs of migration. Adding controls for the costs of migration yields estimates that suggest no significant selection of migrants on observed skill, although estimates are somewhat imprecise. Nevertheless, the substantial change in the pattern of selection when partial controls for the costs of migration are included raises the possibility that analyses with more complete controls would yield evidence

of negative selection. Indeed, analyses that control for the differential in the return to observed skill between Mexico and U.S., which makes the migration decision largely depend only on variation in the costs of migration, indicate that variation in the costs of migration may be the dominant determinant of migration. In this analysis, estimates indicate even stronger evidence of selection from the middle of the observed skill distribution. Theoretically, this is due only to differences in the costs of migration that are correlated with observed skill. Notably, adding controls for the costs of migration to a model that includes differences in the return to observed skills between Mexico and the U.S. again greatly diminishes the evidence in support of selection form the middle of the observed skill distribution. With respect to unobserved skill, adding controls for the costs of migration has little effect on estimates suggesting basically no selection on unobserved skill.

In sum, we presented substantial evidence to support the canonical model of migration. Mexican males are more likely to migrate to the U.S. when the return to observed skill is larger in the U.S. than in Mexico. Moreover, we showed that proxy variables for the cost of migration are strongly correlated with the probability of migrating and correlated with observed skill, and that controlling for these costs diminishes, even eliminates, evidence of selection from the middle of the observed skill distribution. We hypothesize that more complete controls for the costs of migration may yield even more evidence in support of the theoretical prediction of negative selection on skill.

References

Borjas, G.J. (1987) "Self-Selection and the Earnings of Immigrants" *American Economic Review* 77 (4): 531—553

Borjas, G.J. (1991) "Immigration and Self-Selection" in *Immigration, Trade, and the Labor Market* (eds. J.M. Abowd and R.B. Freeman). Chicago: University of Chicago Press (for NBER)

Borjas, G.J., Katz, L.F., and R.B. Freeman (1996) "Searching for the Effect of Immigration on the Labor Market" *American Economic Review* 86(2): 246-251.

Borjas, G.J. (2003) "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market" *Quarterly Journal of Economics* 118(4):1335-1374

Card, D. (2005) "Is the New Immigration Really So Bad?" *Economic Journal* 115(207): 300-323

Chiquiar, D. and G.H. Hanson (2005) "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States" *Journal of Political Economy* 113(2): 239-281

Chiswick, B. (1999) "Are Immigrants Favorably Self-Selected?" *American Economic Review*, 89(2):181-185

Durand, J., Massey, D.S. and R.M. Zenteno (2001) "Mexican Immigration in the United States: Continuities and Changes" *Latin American Research Review* 36: 107–27

Feliciano, C. (2005) "Educational Selectivity in U.S. Immigration: How do Immigrants Compare to Those Left Behind?" *Demography*, 42: 131-152

Fernández-Huertas Moraga, J. (2008) "New Evidence on Emigrant Selection" UFAE and IAE Working Papers 742.08, Unitat de Fonaments de l'Anàlisi Econòmica (UAB) and Institut d'Anàlisi Econòmica (CSIC).

Grogger, J. and G.H. Hanson (2008) "Income Maximization and the Selection and Sorting of International Migrants" NBER Working Paper No. 13821

Hanson, G.H. and C. McIntosh (2010) "The Great Mexican Emigration" *Review of Economics and Statistics*. Forthcoming

Ibarraran, P. and D. Lubotsky (2007) "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census" in *Mexican Immigration to the United States* (eds. G.J. Borjas) University of Chicago Press (for NBER)

LaLonde, R.J., and R.H. Topel (1997) "Economic Impact of International Migration and the Economic Performance of Migrants" in *Handbook of Population and Family Economics*, Vol. 1B, (eds. M.R. Rosenzweig and O. Stark). Amsterdam: Elsevier Science

McKee D. and P.E. Todd (2009) "The Longer-term Effects of Human Capital Enrichment Programs on Poverty and Inequality: Oportunidades in Mexico" mimeo

McKenzie, David and Hillel Rapoport. (2007) "Self-selection patterns in Mexico-U.S. migration: The role of migration networks" *World Bank Policy Research Working Paper* 4118

Munshi, K. (2003) "Networks in the Modern Economy: Mexican Migrants in the US Labor Market" *Quarterly Journal of Economics* 118: 549–99

Orrenius, P.M. and M. Zavodny (2005) "Self-Selection Among Undocumented Immigrants From Mexico," *Journal of Development Economics* 78: 215-240

Passel J. and and D. Cohn (2009), "Mexican Immigrants: How Many Come? How Many Leave?" Pew Hispanic Center Report July 22, 2009

Raven, J.C., Court, J.H., and Raven, J. (1983) *Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 3) - Standard Progressive Matrices*. London: Lewis.

Rubalcava, L.N., Teruel, G.M., Thomas, D. and N. Goldman (2008) "The Healthy Migrant Effect: New Findings From the Mexican Family Life Survey" *American Journal of Public Health* 98(1): 78-84

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

A Data Appendix

This data appendix describes the construction of the main variables:

Migration status: The second round of the Mexican Family Life Survey (MxFLS-2) contains an indicator for individuals who are currently residing in the United States. This is supplemented by information from the migration modules in the second round of the survey on individuals who moved to the United States but who returned to Mexico by time of the second round. This is based on affirmative answers to the questions "Change of residence was to the United States?" for those who indicate migrations of more than 12 months since the first round and "Was the trip to the United States?" for those who indicate moves of less than 12 months since the first round. Return migrants are not excluded because the decision to return to Mexico is likely based on ex-post realizations of their experiences in the United States.

Education: Years of completed schooling is constructed on the basis of three different questions reported directly by individuals in book 3A: (1) "Which is/was the last level of schooling you attended?" with responses including "Without Instruction", "Preschool or Kinder", "Elementary", "Secondary", "Open Secondary", "High school", "Open High school", "Normal basic", "College", or "Graduate"; (2) "Which is the last schooling grade you completed?" ranging from "Did not complete the first grade" to "Seventh grade or more"; and (3) "Did you obtain the degree that certifies you are a graduate at that level?" with responses including "Yes, you graduated/have a degree", "Not yet graduated", and "Have not finished all courses". Question (2) was only asked for those who last attended "Elementary", "Secondary", and "High school" while question (3) was only asked for those who last attended "Open Secondary", "Open High school", "Normal basic", "College", and "Graduate".

Years of completed schooling is taken as 0 for individuals who report "Without Instruction" or "Preschool or Kinder". For individuals who report "Elementary", years of completed schooling ranges from 0 to 6 based on the number of grades completed (there are no responses for more than "sixth grade"). For individuals who report "Secondary", years of completed schooling ranges from 6 to 9 based on the number of grades completed (coding the less than 1% who report more than "third grade" as 9). For individuals who report "High school", years of completed schooling ranges from 9 to 12 based on the number of grades completed (coding the less than 2.5% who report more than "third grade" as 12). For individuals who report "Open Secondary", years of completed schooling is taken as 9 for those who graduated or have a degree, and 7 otherwise. Similarly, for individuals who report "Open High school" or "Normal basic", years of completed schooling is taken as 12 for those who graduated or have a degree, and 10 otherwise. For individuals who report "College", years of completed schooling is taken as 16 for those who graduated or have a degree, and 14 otherwise. For individuals who report "Graduate", years of completed schooling is taken as 18 for those who graduated or have a degree, and 17 otherwise.

Years of completed schooling is also constructed on the basis of the roster reported on individuals by the head of household (or whomever responds to the household interview). However, only questions (1) and (2) were asked through the roster. Consequently, it is not possible to distinguish graduates from non-graduates for those who last attended "Open Secondary", "Open High school", "Normal basic", "College", and "Graduate". Instead, the proportion of graduates derived from the individual reports is used to impute the average years of completed schooling for those reporting these categories, and these are rounded to the nearest integer.

Employment: Information about whether individuals are working is based on four questions from book 3A (1): "What was your main activity last week?"; (2) "During the past week, did you work (or develop any activity that helped the household expenditure), for at least one hour?"; (3)"Did you work in a family business (agricultural or non agricultural) either being paid or not, during the past week?"; and (4) "Do you have a job (or develop any activity that help the household expenditure), but didn't attend it the past week?". Those who report that they "worked or carried out an activity that helped household expenditure" to question (1) or answer affirmatively to any of the follow-up questions are coded as currently working. Those who answer affirmatively to another follow-up question, "In the last 12 months, have you worked (or developed any activity to help household expenditures)?" are added to those currently working and coded as working during the past year. Finally, those who answer affirmatively to the follow-up question "Have you ever worked (or develop any activity that help the household expenditure)?" are added to those who have worked during the past year to form a indicator for having ever worked.

Individuals who work are also asked to report the number of weeks worked in the past year, "What is the total number of weeks that you worked as [main job] in the last year" and the number of hours normally worked per week, "Normally, how many hours do you work as [main job] per week". Employment information is also constructed on the basis of the roster reported on individuals by the head of household (or whomever responds to the household interview). The question is "During the last 12 months did (...) work or develop any activity to help household expenditure" and those who are reported in the affirmative are coded as employed.

Wages: Earnings in the primary and secondary jobs are reported by individuals for both the past month and the past year in book 3A. For employees, rural laborors, agricultural and non-agricultural workers, earnings are reported on the basis of the following questions (1) "How much did you earn last month, since [past month date] until today, for working as [main job]?" and (2) "How much did you earn in the last 12 months, since [date 12 months ago] until today, for working as [main job]?". For annual earnings in a primary job, individuals were asked to provide the detailed amount in the following categories: wages or salary (after taxes), piecework, commissions and tips, extra hours, christmas bonus, additional bonus, vacation premium, profit distribution, meals, housing, transportation, medical benefits, and others. For monthly earnings in a primary job, individuals were asked to provide the detailed amount in the following categories: wages or salary (after taxes), piecework, commissions and tips, extra hours, meals, housing, transportation, medical benefits, and others. For earnings from secondary jobs, individuals were asked to report the total amount. And if individuals did not provide a detailed breakdown for their primary job, they were also simply asked to report the total amount. Monthly and annual earnings are taken as the sum of earnings from the primary and secondary job, based on the detailed earnings or the total earnings depending on which was reported.

For business proprietors, employers, self-employed workers, and peasants who work on their own plot, earnings are reported on the basis of the following questions (1) "How much money did you earn from working as [main job] during the past month, since [past month date] until today?" and (2) "How much money did you earn from working as [main job] during the past 12 months, since [date 12 months ago] until today?". In each case, individuals report gross income/profits and net income/profits. Earnings for business proprietors, employers, self-employed workers, and peasants who work on their own plot, are taken to be equal to net income/profits, unless missing, in which case gross income/profits are used.

For individuals who are working but do not report positive earnings (approximately 20 percent of the sample), we also calculate their imputed earnings based on age, years schooling, marital status, urban status, state fixed effects and earnings information as reported by the head of household (based on the roster question "In the last 12 months, approximately how much did (...) earn or receive from his job or activity to help household expenditure"). We drop outliers that appear to be data errors. Specifically, we trim individuals with hourly wages less than 0.1 and greater than 1000 pesos. This accounts for approximately 0.5 percent of the sample.

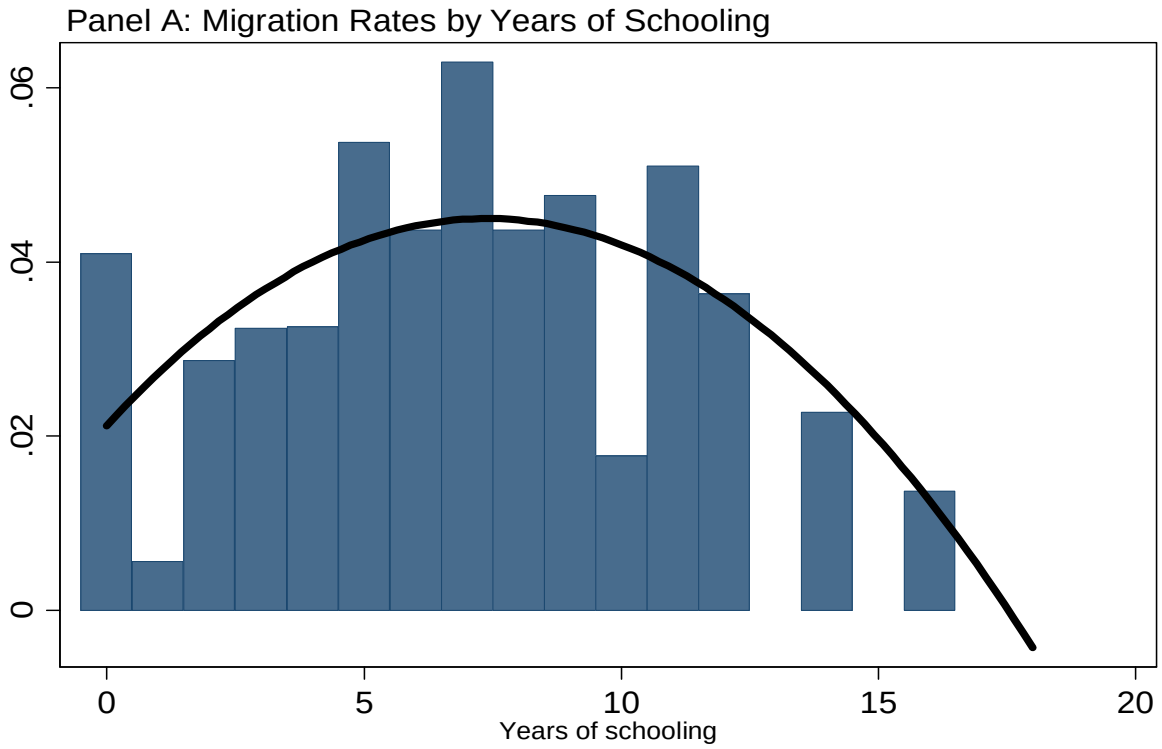
Other variables: Information on marital status and age is available from the roster file. Household assets are taken as the total value of all household asset categories by the head of household, which include dwellings and land, bicycles, motorcycles and automobiles, electronic and kitchen appliances, savings and financial assets, farming equipment, and livestock. Credit and borrowing information is available from individual reports. An indicator for having the ability to borrow is based on an affirmative answer to the question "Do you know any person or place where you can borrow or ask for a credit?". An indicator for having any savings is based on an affirmative answer to the question "Do you have savings?". Cognitive ability is based on the number of correct responses to a Raven's Progressive Matrices test with 12 multiple-choice questions.

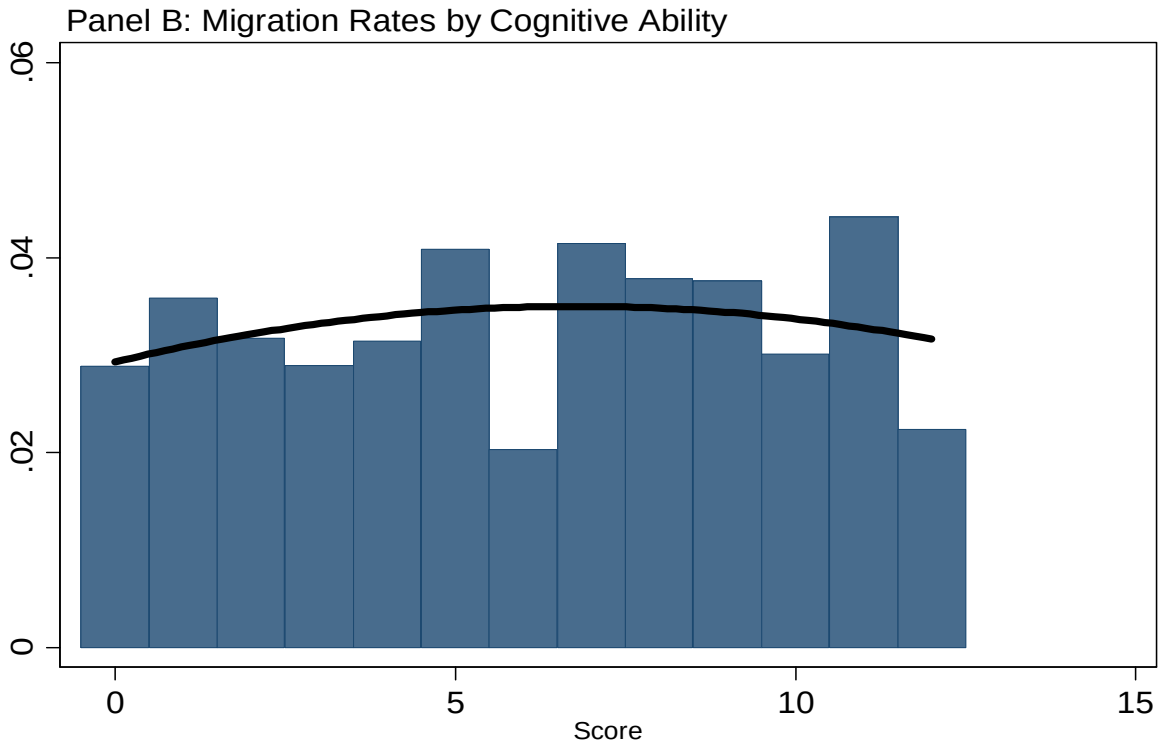
An indicator for having a relative in the US is based on an affirmative answer to the question "Do you have any relative living in the US?". Indicators for whether individuals visited the US in the past (i.e. prior to the first round) is based on affirmative answers to the questions "Change of residence was to the United States?" for those who indicate migrations of more than 12 months and "Was the trip to the United States?" for those who indicate moves of less than 12 months. An indicator for whether individuals have any US documentation is based on those who respond to the question "At the moment of moving to [...], did you have any document that allowed the entry into the USA?" with "Visa", "Green card", or "American citizenship". An indicator for whether individuals have thought about moving to the US is based on those who answer affirmatively to the question "Have you thought about moving in the future, outside the locality/community where you currently live?" and respond with "United States" to the follow-up question "Where do you think

you could move to?".

Proxy information: A proxy book was fielded to collect information about household members that could not be interviewed in person, mainly because they were absent during the time of the fieldwork. Information in the proxy is based on a more limited set of identical questions to those presented in the individual one-on-one interviews, and provided by a household member knowledgeable about the absent member. Proxy reports account for approximately 10-15 percent of observations for most variables.

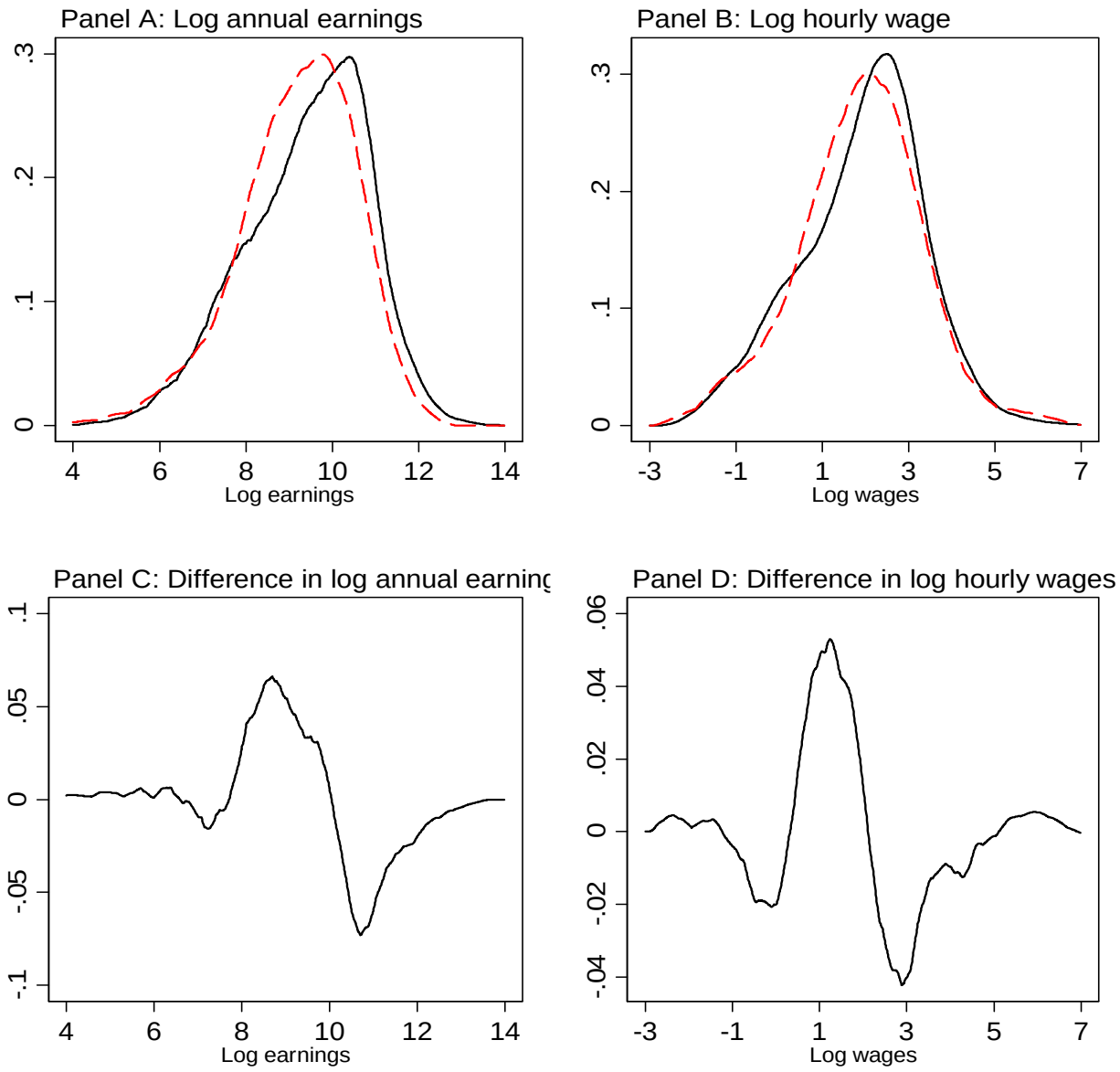
Figure 1: Migration Rates by Education and Cognitive Ability





Notes: Panel A plots migration rates for Mexican males aged 21 to 65 by years of schooling, with the solid line showing the quadratic relationship between the probability of migrating (estimated at the individual level) and measure of skill. Panel B plots migration rates for Mexican males aged 21 to 65 by cogntive ability score, with the solid line showing the analogous quadratic relationship. Source: Mexican Family Life Survey

Figure 2: Wage Distributions by Migration Status



Notes: Panel A plots the density distribution of log annual earnings for migrants (dashed) and non-migrants (solid). Panel B plots the density distribution of log hourly wages for migrants (dashed) and non-migrants (solid). Panel C plots the difference in the densities of migrant and non-migrant annual earnings. Panel D plots the difference in the densities of migrant and non-migrant hourly wages. All densities are estimated using an Epanechnikov kernel and Silverman's (1986) optimal bandwidth.
Source: Authors calculations from the Mexican Family Life Survey

Table 1: Summary Statistics

	Mean	Standard deviation	Observations
Panel A: Migration information			
Migrated to US between 2002-2005	0.036	0.187	8,116
Panel B: Roster reports			
Age	38.9	12.222	8,116
Married	0.63	0.484	8,113
Rural	0.49	0.500	8,114
Years of schooling	7.29	4.192	8,057
Worked	0.911	0.284	8,114
Annual earnings	37,461	45,276	6,014
Household assets	751,680	10,400,000	7,530
Panel C: Individual reports			
Years of schooling	7.31	4.484	7,223
Cognitive ability (Raven's Progessive Matrices)	5.894	3.013	6,297
Relative in US	0.348	0.476	6,261
Visited US	0.043	0.203	7,231
Has US documents	0.014	0.117	6,234
Thought of moving to US	0.028	0.164	6,226
Know person/places to borrow	0.406	0.491	7,240
Have savings	0.157	0.364	7,240
Worked	0.901	0.299	7,221
Worked during past year	0.941	0.236	7,221
Ever worked	0.972	0.166	7,221
Total annual hours	2238.5	998.428	6,178
Annual earnings	28,192	42,201	5,924
Hourly wage	17.01	37.631	5,613

Notes: Information in Panel A is based on an indicator for whether the individual has moved to the United States, together with information on men who moved to the United States but returned to Mexico by time of the second round. Panel B contains information from the first round of the MxFLS and reported by the head of household (or other knowledge person in the household). Panel C contains information from the first round of the MxFLS and is reported by the individual directly, using proxy reports to complete any missing information. See Data Appendix for more details on the construction of these variable. All summary statistics are calculated for Mexican males aged 21 to 65, and are unweighted. Source: Mexican Family Life Survey.

Table 2: Selected Characteristics by Migration Status

	Migrants	Non-migrants	Difference
Age			
21-29 years old	0.441	0.277	0.163 ***
30-38 years old	0.231	0.250	-0.020
39-47 years old	0.207	0.205	0.002
48-56 years old	0.078	0.157	-0.079 ***
57-65 years old	0.044	0.110	-0.066 ***
Education			
0-3 years of schooling	0.202	0.239	-0.037
4-6 years of schooling	0.301	0.262	0.040
7-9 years of schooling	0.320	0.246	0.073 ***
10-12 years of schooling	0.121	0.127	-0.006
>12 years of schooling	0.055	0.126	-0.071 ***
Cognitive ability			
Bottom quartile	0.151	0.155	-0.005
2nd quartile	0.307	0.306	0.000
3rd quartile	0.307	0.314	-0.007
Top quartile	0.236	0.224	0.012
Annual earnings			
1st (bottom) quintile	0.198	0.200	-0.002
2nd quintile	0.252	0.192	0.060 *
3rd quintile	0.238	0.204	0.033
4th quintile	0.198	0.200	-0.002
5th (top) quintile	0.114	0.204	-0.090 ***
Hourly wages			
1st (bottom) quintile	0.188	0.200	-0.013
2nd quintile	0.255	0.198	0.057 **
3rd quintile	0.188	0.200	-0.012
4th quintile	0.188	0.201	-0.014
5th (top) quintile	0.182	0.201	-0.018
Other variables			
Worked	0.882	0.902	-0.019
Married	0.573	0.628	-0.055 *
Relative in US	0.551	0.341	0.209 ***
Visited US	0.165	0.038	0.127 ***
Has US documents	0.039	0.013	0.026 ***
Thought of moving to US	0.092	0.026	0.067 ***

Notes: ***, ** and * indicate statistically significant differences between migrants and non-migrants at the 1, 5 and 10 percent level respectively. The sample is Mexican males aged 21 to 65 and unweighted. Source: Mexican Family Life Survey.

Table 3: The Association between Predicted Earnings/Wages and Migration

dependent variable: migrated to the US	Predicted Annual Earnings			Predicted Hourly Wage		
	(1)	(2)	(3)	(4)	(5)	(6)
2nd Quintile	0.007	0.009	0.008	0.006	0.017**	0.003
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
3rd Quintile	0.013*	0.026***	0.016**	0.013*	0.031***	0.015*
	[0.007]	[0.008]	[0.008]	[0.007]	[0.008]	[0.008]
4th Quintile	-0.001	0.017**	0.018**	0	0.024***	0.023***
	[0.008]	[0.007]	[0.008]	[0.008]	[0.007]	[0.008]
5th Quintile	-0.018***	-0.005	0.003	-0.018***	-0.001	0
	[0.007]	[0.006]	[0.007]	[0.007]	[0.006]	[0.007]
Predicting variables:						
Education	X	X	X	X	X	X
Education*Age		X	X		X	X
Education*Age*Ability			X			X
Sample size	5826	5826	5826	5608	5608	5608
Mean of dep. var.	0.0346	0.0346	0.0346	0.0340	0.0340	0.0340

Notes: Robust standard errors clustered at the household level are in brackets. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. The dependent variable is equal to 1 if migrated to the United States between 2002 and 2006, and 0 otherwise. The independent variables are quintiles of predicted log earnings and wages. These are constructed by regressing annual earnings or hourly wages on 5 education categories (in columns 1 and 4), 5 education and 5 age categories, fully interacted (in columns 2 and 5), and 5 education, 5 age, and 4 cognitive ability categories, fully interacted (in columns 3 and 6). The omitted quintile is the 1st (bottom) quintile of predicted earnings or wages. All regressions are for Mexican males aged 21 to 65, and are unweighted. Source: Mexican Family Life Survey.

Table 4: The Association between Residual Earnings/Wages and Migration

dependent variable: migrated to the US	Residual Annual Earnings			Residual Hourly Wage		
	(1)	(2)	(3)	(4)	(5)	(6)
2nd Quintile	0.004	0.003	-0.003	0.006	0.005	0.004
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
3rd Quintile	-0.002	-0.006	-0.006	0.01	0.007	0
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
4th Quintile	-0.003	-0.006	-0.01	0.004	0	-0.004
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
5th Quintile	-0.016**	-0.015*	-0.021***	0	-0.002	-0.004
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]
Residualizing variables:						
Education	X	X	X	X	X	X
Education*Age		X	X		X	X
Education*Age*Ability			X			X
Sample size	5,826	5,826	5,826	5,826	5,826	5,826
Mean of dep. var.	0.0346	0.0346	0.0346	0.0340	0.0340	0.0340

Notes: Robust standard errors clustered at the household level are in brackets. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. The dependent variable is equal to 1 if migrated to the United States between 2002 and 2006, and 0 otherwise. The independent variables are quintiles of residual log earnings and wages. These are constructed by regressing annual earnings or hourly wages on 5 education categories (in columns 1 and 4), 5 education and 5 age categories, fully interacted (in columns 2 and 5), and 5 education, 5 age, and 4 cognitive ability categories, fully interacted (in columns 3 and 6). The omitted quintile is the 1st (bottom) quintile of predicted earnings or wages. All regressions are for Mexican males aged 21 to 65, and are unweighted. Source: Mexican Family Life Survey.

Table 5: The Association between Predicted Earnings/Wages and Migration with Additional Controls

dependent variable: migrated to the US

Panel A: Predicted Annual Earnings

	(1)	(2)	(3)	(4)	(5)	(6)
2nd Quintile	0.008	0.008	0.008	0.019**	0.019	0.013
	[0.007]	[0.007]	[0.007]	[0.007]	[0.012]	[0.011]
3rd Quintile	0.016**	0.011	0.018**	0.033***	0.038***	0.021*
	[0.008]	[0.008]	[0.008]	[0.008]	[0.011]	[0.010]
4th Quintile	0.018**	0.009	0.020***	0.040***	0.045**	0.023
	[0.008]	[0.007]	[0.008]	[0.008]	[0.018]	[0.017]
5th Quintile	0.003	-0.008	0.007	0.031***	0.053**	0.018
	[0.007]	[0.007]	[0.007]	[0.008]	[0.025]	[0.024]
Differential returns					0.053**	0.026
					[0.019]	[0.018]
"Cost" controls		X				X
"Credit" controls			X			
Community fixed effects				X		
Sample size	5,826	5,826	5,826	5,826	5,826	5,826

Panel B: Predicted Hourly Wage

	(1)	(2)	(3)	(4)	(5)	(6)
2nd Quintile	0.003	0.002	0.004	0.012	0.015	0.009
	[0.007]	[0.007]	[0.007]	[0.007]	[0.010]	[0.009]
3rd Quintile	0.015*	0.01	0.017**	0.028***	0.036***	0.022**
	[0.008]	[0.008]	[0.008]	[0.008]	[0.011]	[0.010]
4th Quintile	0.023***	0.011	0.025***	0.041***	0.048***	0.027*
	[0.008]	[0.008]	[0.008]	[0.009]	[0.016]	[0.014]
5th Quintile	0	-0.009	0.004	0.025***	0.052**	0.022
	[0.007]	[0.007]	[0.007]	[0.008]	[0.024]	[0.022]
					[0.020]	[0.018]
"Cost" controls		X				X
"Credit" controls			X			
Community fixed effects				X		
Sample size	5,608	5,608	5,608	5,608	5,608	5,608

Notes: Robust standard errors clustered at the household level are in brackets. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. The dependent variable is equal to 1 if migrated to the United States between 2002 and 2006, and 0 otherwise. The independent variables are quintiles of predicted log earnings and wages, constructed by regressing annual earnings or hourly wages on 5 education, 5 age, and 4 cognitive ability categories, fully interacted. Column 2 "cost" controls include: marital status, relative in the U.S., visited the U.S., having U.S. documentation, and thought about moving to the U.S. Column 3 "credit" controls include: household assets, ability to borrow, having savings. Column 4 adds 150 community level fixed effects. Column 5 includes a variable capturing differential returns between Mexico and the U.S. by 25 education*age categories, and is clustered on these education*age categories. All regressions are for Mexican males aged 21 to 65, and are unweighted. Source: Mexican Family Life Survey.

Table 6: The Assocation between Residual Earnings/Wages and Migration with Additional Controls

dependent variable: migrated to the US

Panel A: Residualized Annual Earnings

	(1)	(2)	(3)	(4)
2nd Quintile	-0.003	-0.006	-0.003	-0.002
	[0.008]	[0.008]	[0.008]	[0.008]
3rd Quintile	-0.006	-0.009	-0.006	-0.001
	[0.008]	[0.008]	[0.008]	[0.008]
4th Quintile	-0.01	-0.011	-0.01	0.001
	[0.008]	[0.008]	[0.008]	[0.008]
5th Quintile	-0.021***	-0.020***	-0.020***	-0.009
	[0.007]	[0.007]	[0.008]	[0.008]
"Cost" controls		X		
"Credit" controls			X	
Community fixed effects				X
Sample size	5,826	5,826	5,826	5,826

Panel B: Predicted Hourly Wage

	(1)	(2)	(3)	(4)
2nd Quintile	0.004	0.001	0.004	0.01
	[0.008]	[0.008]	[0.008]	[0.008]
3rd Quintile	0	-0.004	0	0.011
	[0.008]	[0.008]	[0.008]	[0.008]
4th Quintile	-0.004	-0.006	-0.003	0.008
	[0.008]	[0.008]	[0.008]	[0.008]
5th Quintile	-0.004	-0.006	-0.002	0.007
	[0.008]	[0.008]	[0.008]	[0.008]
"Cost" controls		X		
"Credit" controls			X	
Community fixed effects				X
Sample size	5,608	5,608	5,608	5,608

Notes: Robust standard errors clustered at the household level are in brackets. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. The dependent variable is equal to 1 if migrated to the United States between 2002 and 2006, and 0 otherwise. The independent variables are quintiles of residual log earnings and wages, constructed by regressing annual earnings or hourly wages on 5 education, 5 age, and 4 cognitive ability categories, fully interacted. Column 2 "cost" controls include: marital status, relative in the U.S., visited the U.S., having U.S. documentation, and thought about moving to the U.S. Column 3 "credit" controls include: household assets, ability to borrow, having savings. Column 4 adds 150 community level fixed effects. All regressions are for Mexican males aged 21 to 65, and are unweighted. Source: Mexican Family Life Survey.

Appendix Table 1: Robustness checks for Predicted Earnings/Wages

dependent variable: migrated to the US

Panel A: Predicted Earnings

	Monthly earnings		No-imputed annual		No-proxied annual	
	(1)	(2)	(3)	(4)	(5)	(6)
2nd Quintile	0.012	0.009	0.006	0.006	0	-0.002
	[0.008]	[0.008]	[0.008]	[0.008]	[0.007]	[0.007]
3rd Quintile	0.009	0.006	0.019**	0.014*	0.020**	0.018**
	[0.008]	[0.008]	[0.008]	[0.008]	[0.009]	[0.009]
4th Quintile	0.001	0	0.014*	0.005	0.005	0.001
	[0.007]	[0.007]	[0.008]	[0.008]	[0.007]	[0.007]
5th Quintile	-0.009	-0.012*	0.001	-0.01	-0.01	-0.011*
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.006]
"Cost" controls		X		X		X
Sample size	5819	5819	4661	4661	5093	5093

Panel B: Predicted Hourly wages

	Monthly-based hourly		No-imputed hourly wages		No-proxied hourly wages	
	(1)	(2)	(3)	(4)	(5)	(6)
2nd Quintile	0.012	0.009	0	-0.002	0	-0.001
	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]	[0.007]
3rd Quintile	0.009	0.006	0.018**	0.013	0.020**	0.015*
	[0.008]	[0.008]	[0.009]	[0.009]	[0.008]	[0.008]
4th Quintile	0.013*	0.01	0.016*	0.006	0.016**	0.012
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
5th Quintile	-0.011*	-0.013*	0.001	-0.009	-0.011*	-0.013**
	[0.007]	[0.007]	[0.008]	[0.007]	[0.006]	[0.006]
"Cost" controls		X		X		X
Sample size	5606	5606	4526	4526	4947	4947

Notes: Robust standard errors clustered at the household level are in brackets. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. The dependent variable is equal to 1 if migrated to the United States between 2002 and 2006, and 0 otherwise. The independent variables are quintiles of predicted and residual log earnings and wages. Columns 2, 4, and 6 include controls for marital status, relative in the U.S., visited the U.S., having U.S. documentation, and thought about moving to the U.S. All regressions are for Mexican males aged 21 to 65, and are unweighted. Source: Mexican Family Life Survey.