

New Evidence on Emigrant Selection *

Jesús Fernández-Huertas Moraga[†]

April 10, 2008

Abstract

This paper examines the extent to which Mexican emigrants to the United States are *negatively selected*, that is, have lower skills than individuals who remain in Mexico. Previous studies have been limited by the lack of nationally representative longitudinal data. This one uses a newly available household survey, which identifies emigrants before they leave and allows a direct comparison to non-migrants. I find that, on average, US bound Mexican emigrants from 2000 to 2004 earn a lower wage and have less schooling years than individuals who remain in Mexico, evidence of *negative selection*. This supports the original hypothesis of Borjas (AER, 1987) and argues against recent findings, notably those of Chiquiar and Hanson (JPE, 2005). The discrepancy with the latter is primarily due to an under-count of unskilled migrants in US sources and secondarily to the omission of unobservables in their methodology.

Keywords: international migration, selection, household survey

JEL Classification Numbers: F22, O15, J61, D33

*I am particularly thankful to Ronald Findlay, Eric Verhoogen and David Weinstein for their help and support. This paper has also benefited from useful comments and suggestions from Claudia Canals, Ernesto Castañeda, Pierre-André Chiappori, Barry Chiswick, Janet Currie, Donald Davis, Chico Ferreira, Olga Gorbachev, Gordon Hanson, James Harrigan, Guillermina Jasso, Wojciech Kopczuk, Anton Korinek, David S. Lee, Tomasz Michalski, Camelia Minoiu, Kiki Pop-Eleches, Francisco Rivera-Batiz, Johannes Schmieder, Miguel Urquiola and seminar participants at Columbia University, Universidad de Oviedo, Universidad de Alicante, Syracuse University, Tilburg University, IESE, the Instituto de Análisis Económico (IAE-CSIC) and Universitat Pompeu Fabra. Of course, remaining errors are only mine.

[†]Instituto de Análisis Económico (CSIC), Campus UAB, 08193 Bellaterra (Barcelona), Spain. E-mail: jesus.fernandez@iae.csic.es

1 Introduction

This paper examines how the productive characteristics of Mexican emigrants to the United States compare to non-migrants. The extent of emigrant selection is important because it affects the level and the distribution of welfare in both emigrant-sending and immigrant-receiving countries.¹ Thus, economists have long tried to explain how emigrants self select. On the theory side, Borjas (1987) stated that most immigrants should be low skilled when the reward to skills or earnings inequality in their home country is higher than the reward to skill or earnings inequality in the receiving country.² This is the case between Mexico and the United States (Hanson (2006)) so that negative selection, meaning that emigrants are on average less skilled or productive than those who do not migrate, should characterize migration flows between the two countries. However, Chiquiar and Hanson (2005) found that Mexican immigrants in the United States originated in the medium-high range of the Mexican wage distribution, which is interpreted as evidence of positive selection. Their data on Mexican immigrants came mainly from the US Census, which is known to suffer from under-counting immigrants, especially if they are undocumented.³ Other studies⁴ finding positive selection were mainly based on the Mexican Migration Project, which provides detailed information about a particular region in Mexico but it is not representative of the whole country.⁵

¹The magnitude of the impact of migration flows on welfare is controversial. On the effects of general migration flows on welfare, see World Bank (2006). On the effect of immigration into the United States on US wages and welfare, see Borjas (2003) and Ottaviano and Peri (2006). In the case of Mexican immigration to the United States, see Borjas and Katz (2007). Finally, on the effects of emigration from Mexico on Mexican wages and welfare, see Mishra (2007).

²The reason is that he assumes that migration costs are constant in the skill level of individuals so that those with low skills have relatively more incentives to migrate than those who have high skills that are better rewarded in their home country.

³Hanson (2006) reviews some of the estimates of the under-count of undocumented migrants in the US Census and they range from 10 to 25 per cent of the emigrant stock. In addition, the problem is likely to be more severe for recently arrived immigrants.

⁴McKenzie and Rapoport (2007a) and Orrenius and Zavodny (2005), for example.

⁵The MMP is an ongoing research project between Princeton University and the Universidad de Guadalajara in Mexico. They have been interviewing households in rural traditional migrant sending communities since 1982. The states included in the MMP (Munshi (2003)) represented approximately one quarter of total Mexican population and one half of total migration flows from Mexico for the period 1997-2002 (INEGI (2004)). The level of representation of the MMP is smaller than this since it only covers selected rural

This paper addresses these data problems by using a newly available data source: the Encuesta Nacional de Empleo Trimestral (ENET) or Quarterly National Labor Survey (INEGI (2005)), a household survey run by the Instituto Nacional de Estadística, Geografía e Informática (INEGI), the Mexican National Statistical Agency, from 2000 to 2004.⁶ The ENET structure is similar to that of the Current Population Survey (CPS) in the United States. It is representative at a national level and it follows households for five quarters so that it allows recovering the wage income and other characteristics (like education) of Mexican emigrants (both documented and undocumented) in the previous quarter to that in which they decided to leave the country.

Throughout this paper, a Mexican emigrant to the United States is defined as an individual who is in Mexico at the time her wage (and other characteristics) is observed but leaves for the United States in the following quarter, when her household is surveyed again. A non-migrant is defined analogously as an individual who is in Mexico at the time her wage is observed and who remains in Mexico in the following quarter, when her household is surveyed again.

The wage distribution of Mexican emigrants is shifted to the left of the wage distribution of Mexican non-migrants for both men and women. As long as the wage is a valid measure of the marginal product of labor, this implies that Mexican emigrants to the United States are on average less productive than those who remain at home. In conclusion, the main result of this paper is the existence of *negative selection* in the emigration flows from Mexico to the United States for the period 2000-2004.

The procedure to test the selectivity of emigrants from Mexico to the United States is similar to the methodology in Chiquiar and Hanson (2005). However, the ENET, which was not available at the time they wrote their paper, does not require the construction of counterfactual wage densities since actual wages of future emigrants can be observed directly. The only exercise is to estimate wage densities for migrants and non-migrants based on the direct observation of their wages at the same period and to compare them. Contrary to them, I find evidence of negative selection in terms of skills (both observable and unobservable)

communities in these states. For more information on the MMP, visit <http://mmp.opr.princeton.edu/>

⁶The ENET was created by combining two previously existing surveys: the ENE (Encuesta Nacional de Empleo) and the ENEU (Encuesta Nacional de Empleo Urbano). See below for details. After 2004, the ENET was substituted by the Encuesta Nacional de Ocupación y Empleo (ENOE). I thank Eric Verhoogen for facilitating my access to the ENET dataset.

reflected in the wage levels. The comparison of the schooling distributions of migrants and non-migrants also displays negative selection.

The discrepancy between this paper and Chiquiar and Hanson (2005) could be related to three factors. First, their results reflect selection in the stock of migrants in the United States, obtained from the US Census, whereas the information in the ENET corresponds to selection on the flow of migrants. Second, their methodology cannot take potentially relevant unobservable characteristics into consideration in the estimation of counterfactual wages. Third, US sources are known to under-count Mexican immigrants (see footnote 3). These explanations can be explored with the ENET data. When performing this exercise, I find evidence that selection is negative both on observable and on unobservable characteristics. These results suggest that the under-count of low-skilled immigrants in US data sources is the main reason why this paper finds negative selection whereas Chiquiar and Hanson (2005) found positive selection.

Self selection of migrants has traditionally been one of the most controversial topics in the migration literature. Chiswick (1978) claimed that only the more capable individuals would have the drive and motivation to locate in a new country, which would lead to positive selection of emigrants. In more recent papers, Chiswick (1999) and Chiswick (2007), using the human capital migration model, shows that positive selection in the supply of migrants is to be expected as long as the earnings structure of the country of destination coincides with that of the country of origin. On the contrary, Borjas (1987) showed that, in cases where the return to skills is higher in the origin country than in the destination country, low skill individuals have more incentives to emigrate, with a resulting migrant pool which becomes negatively selected.⁷ However, simple extensions of Borjas' model show that both positive and negative selection of immigrants are theoretically possible even when the reward to skills is higher in the origin country. Studying the composition of migration flows or "Who are the immigrants?" (Borjas (1990)) becomes thus an empirical question.

There is a long tradition of testing this empirical question by concentrating on migration flows between Mexico and the United States. In addition to Chiquiar and Hanson (2005), McKenzie and Rapoport (2007a) and Orrenius and Zavodny (2005) have found that the probability of emigration of Mexican individuals increases in the upper and middle sections

⁷This possibility is also addressed by Chiswick (1999) and Chiswick (2007) in the context of the human capital model.

of the schooling distribution, which is also evidence of positive selection. Both studies used data from the MMP, which, as mentioned above, is focused on a limited geographic area (see footnote 5). I argue below that the MMP is not representative of total Mexican emigration and offers a biased picture if used to infer general results. By contrast, Ibararán and Lubotsky (2007) used the Mexican Census to estimate the schooling level of Mexican emigrants from the schooling level of family members left behind and found it was lower than the schooling level of non-migrants.⁸ Finally, McKenzie and Rapoport (2007b) suggest that all of these conflicting results can be reconciled by taking into account the role of migration networks. They find that selection is positive in areas with low migration networks and negative in areas with high migration networks. The reason is that networks reduce the costs and increase the benefits of the migratory move.⁹

The structure of the paper follows. First, the economic theory underlying this study is sketched. Second, a brief description of the ENET dataset is presented. Third, the main result of the paper (negative selection) is developed. A fourth section assesses the source of the differences with previous literature and why the results in this paper are still consistent with that previous literature. Finally, the main conclusions are drawn.

2 Emigrant Selection Theory

The migration move is traditionally modeled as an investment decision, following Sjaastad (1962). Every individual assesses the expected utility to be obtained in each possible destination and decides to locate wherever this expected utility is higher at every point in time. In this framework, the migration investment will be preferred to alternative investments

⁸Cuecuecha (2005) tries to reconcile both points of view (positive versus negative selection) by using more sophisticated choice-based techniques when combining the US and Mexican sources. Using the same sources (US and Mexico 2000 Census), Caponi (2006) finds however that there is a U-shape relationship between the emigration probability and the level of education, with most migrants thus coming from the bottom and the top of the schooling distribution in Mexico.

⁹McKenzie and Rapoport (2007b) estimate their model from the Encuesta Nacional de la Dinámica Demográfica (ENADID) for 1997. This is a special survey that the INEGI developed in 1992 and 1997 with detailed nationally representative information about migrant behavior. However, McKenzie and Rapoport (2007b) restrict their analysis to individuals living in localities with less than 100,000 inhabitants (around 49 per cent of the population in 1995), since they claim the effects of the network are more difficult to realize for people living in highly urbanized areas.

(education, starting a firm...) whenever the expected return to migrating is higher than other expected returns. Emigrants will be those whose return to migrating is higher due to both their observable and unobservable characteristics. The differing characteristics between migrants and non-migrants are the object of study: do emigrants have more productive characteristics than non-migrants? Yes implies positive selection. No means negative selection. There can be either positive or negative selection of emigrants depending on the parameters of the model.

Following Borjas (1999) rendition of Roy (1951) model, consider a set of individuals living at location 0 (Mexico) who must decide whether to move to location 1 (United States) or else remain at 0. Suppose that individuals can be characterized by a vector of observable and unobservable characteristics that will determine their performance in the labor market. To simplify things further, assume that these characteristics can be summarized in the variable $x \geq 0$, whose distribution function over the population is $F(x)$. The logarithm of the wage at location 0 (w_0) is:

$$\log w_0 = \mu_0 + \delta_0 x$$

where $\mu_0 > 0$ can be interpreted as the base wage and $\delta_0 > 0$ is the return to personal characteristics at location 0. If an individual with characteristics x decides to migrate to location 1, the alternative wage that can be obtained there will be given by:

$$\log w_1 = \mu_1 + \delta_1 x$$

Assume that $\mu_1 > \mu_0$ and $\delta_1 < \delta_0$, that is, the base wage is higher at location 1 (United States) whereas the return to personal characteristics (suppose, to fix ideas, that the most important of these personal characteristics is education) is higher at location 0 (Mexico). Also, $\delta_1 > 0$, which means that whatever is considered productive in Mexico, it is also considered productive in the United States.¹⁰

An income maximizing individual will migrate whenever the wage in the destination net of migration costs ($C > 0$) exceeds the wage at her original location. This can be expressed with the following function:

¹⁰This corresponds to Borjas (1999) assumption that the correlation coefficient between earnings distributions in both locations is positive.

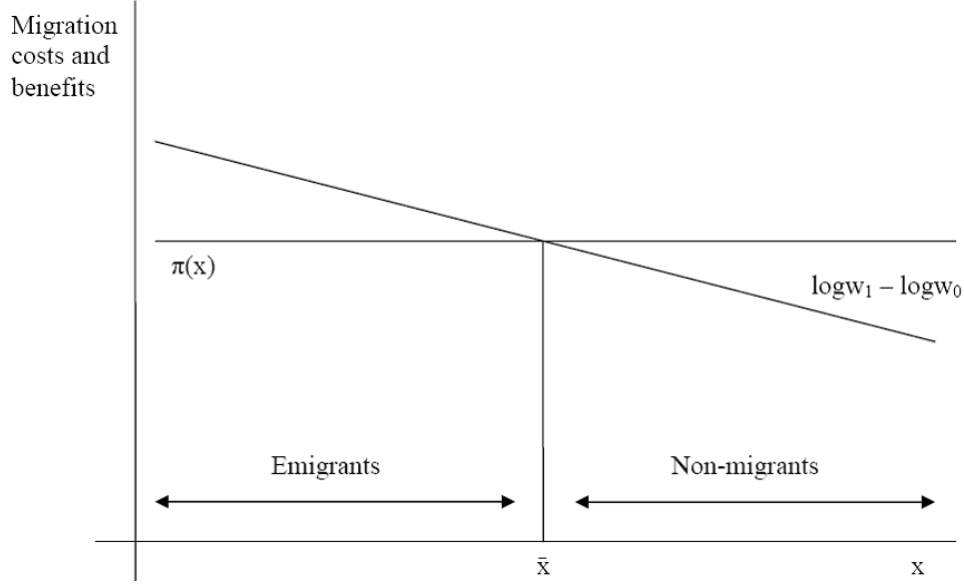


Figure 1: Negative selection for migration costs constant in productive characteristics (x)

$$I(x) \equiv \log \left(\frac{w_1}{w_0 + C} \right) \simeq \log w_1 - \log w_0 - \pi > 0$$

where $\pi = \frac{C}{w_0}$ are migration costs in time-equivalent units. If π are considered constant across characteristics, only individuals with characteristics below \bar{x} will migrate, with

$$\bar{x} = \frac{\mu_1 - \mu_0 - \pi}{\delta_0 - \delta_1}$$

This will imply negative selection of emigrants. The less productive individuals will migrate whereas the more productive individuals will decide to remain at 0 (Borjas (1987)). This situation is depicted in figure 1.

Next, Borjas (1999) definition of positive selection¹¹ can be introduced to this simpler setup as a situation in which:

$$E(\log w_0 | I(x) > 0) > E(\log w_0 | I(x) < 0)$$

In words, positive selection implies that emigrants are on average more productive (as reflected on their wage) than non-migrants. The above inequality can be easily computed

¹¹Borjas (1999) definition actually also includes that the earnings of immigrants will be higher than those of natives in the host country as long as the base average wage both groups have access to is the same.

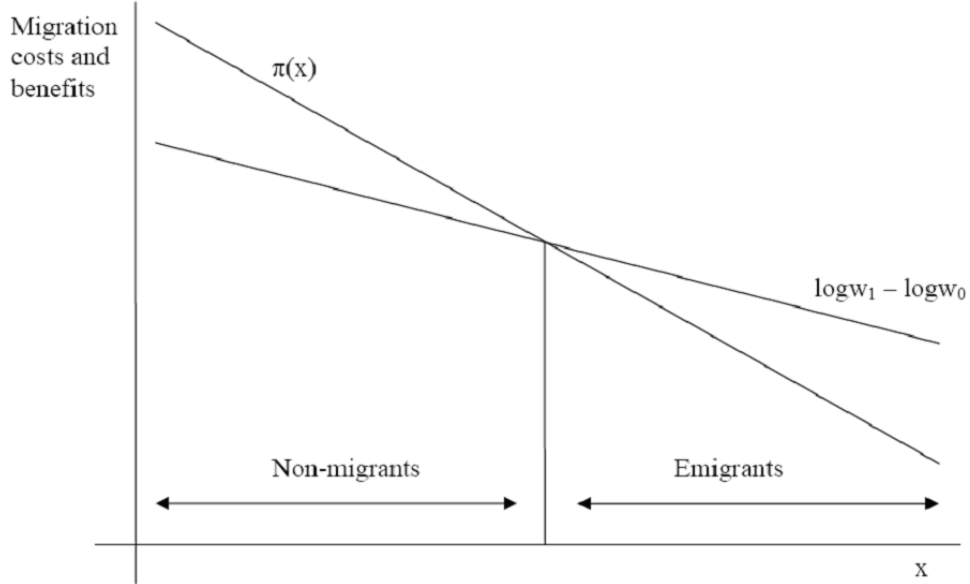


Figure 2: Positive selection for migration costs decreasing in productive characteristics (x)

from the ENET data for the Mexico-US case since both the wages of non-migrants and migrants right before migration can be observed. In addition, the difference between the two expectations can be interpreted as the degree of selection (DS):

$$DS \equiv E(\log w_0 | I(x) > 0) - E(\log w_0 | I(x) < 0)$$

Since $\mu_1 > \mu_0$ and $\delta_1 < \delta_0$, DS will be negative unless migration costs are so large that $\bar{x} < 0$ so that nobody migrates. The degree of selection is increasing in \bar{x} , which is increasing in the base wage (μ_1) and return to skills (δ_1) at location 1 and decreasing in both the base wage (μ_0) and return to skills (δ_0) at location 0 and also in migration costs (π). Thus, clear predictions about the evolution of selection and its relationship with basic variables can be extracted from this simple model.

Notice that a constant π means that real migration costs are actually increasing in x . In general, any model with π increasing in x will also deliver the same predictions about migrant selectivity.

Chiquiar and Hanson (2005) show that assuming instead that π is decreasing in x leads to the possibility of both negative or positive selection depending on parameter values. Figure 2 depicts a case in which selection is positive.

The structure of migration costs can give rise to many different migration patterns characterized by positive, negative or intermediate selection. A priori, the relationship between productive characteristics and migration costs can be argued to go in both directions (Ibararán and Lubotsky (2007)). More productive individuals may decide to migrate legally to be able to enjoy high returns to their characteristics in the destination country (Hanson (2006)). Migrating legally usually requires longer waiting times with the corresponding higher costs so that we would observe a positive relationship between migration costs and skill levels. In general, endogenizing migration costs will lead to this positive relationship even for illegal migrants since they may decide to spend more on better illegal crossing strategies (Gathmann (2004)) or even on their traveling.

On the other hand, since the relevant concept of costs refers to time-equivalent units, it is obvious that the same level of real costs becomes more onerous for low wage individuals. There is another case in which migration costs end up being decreasing in productive characteristics: the case of credit or wealth constraints. An individual is constrained in wealth when she would be willing to migrate given her expected return to migration ($I(x) > 0$) but she cannot afford the trip. If credit markets worked efficiently, this individual should be able to borrow in order to undertake migration.

The general point that can be established from the study of emigrant selection theory is that many different and complex selection patterns can emerge from very simple assumptions so that determining how emigrants end up self selecting is primarily an empirical question.

3 The ENET Dataset

The main source used in this paper is the ENET. The ENET is the household survey the INEGI used to calculate the official employment statistics for Mexico from the second quarter of 2000 until the end of 2004.¹² Before 2000, the only labor survey at the national level was the Encuesta Nacional de Empleo (ENE) but it was only carried out yearly so that the kind of analysis used here cannot be applied to those data. However, quarterly series for the

¹²After that, the ENET was substituted by a new survey: Encuesta Nacional de Ocupación y Empleo (ENOE).

urban areas are available since 1987 in the Encuesta Nacional de Empleo Urbano (ENEU).¹³

The ENET is very similar to the Current Population Survey in the United States. It interviews households and it gathers information about all of their members, including whether any of them migrated to the United States. Since every household is interviewed five times, with one of the five panels dropping out of the sample each quarter, a researcher can match the data on wages or schooling of an individual in a quarter in which she lives in Mexico with the migration behavior of that individual in the following quarter.¹⁴

Some definitions must be established at this point. Throughout the paper, emigrants are individuals aged 16 to 65 years old¹⁵ who are present in Mexico at quarter t and who are reported to have left for the United States when the interviewer returns to the same household at quarter $t+1$. Conversely, non-migrants are individuals from the same age group who are present in Mexico at quarter t and who are reported to live still in Mexico (in the same or in a different household) when the interviewer returns to the same household at quarter $t+1$.

The main summary statistics¹⁶ for the ENET in the studied period are presented in table 1, comparing data on migrants with data on non-migrants for both men and women.

Mexican migrants earn on average a significantly lower wage than non-migrants, both for men and women. If wages reflect productive capabilities, it can be concluded that male emigrants to the United States are approximately 28 per cent less productive than those who stay in Mexico for the 2000-2004 period, which means that the degree of selection is negative. There is also negative selection for women since emigrant women earn on average 77 per cent of what a typical non-migrant has as her wage. Thus, the negative selection result to be developed in the next section already appears in the summary statistics of the data.

¹³For example, Robertson (2000) uses the ENEU to study labor market integration between Mexico and the US.

¹⁴Sample attrition amounts to 8-10 per cent for most of the quarters but it doubles in 2003. The results of the paper are robust to the exclusion of this year. Also, the distribution of wages and characteristics of non-respondents are statistically equivalent to those of respondents so there does not seem to be scope for significant attrition bias.

¹⁵Expanding the sample to those aged 12 to 15 and over 65 years old to take advantage from all the information the ENET provides does not alter the results.

¹⁶All standard errors and estimation commands in this paper are calculated using Stata's linearized svy option, using the survey weights and clustering at the household level.

	<i>Non-Migrants</i>	<i>Migrants</i>	<i>Non-Migrants</i>	<i>Migrants</i>
<i>Observations</i>	4,781,283	13,091		
<i>Percent Male</i>	47%	81%		
	(0.00)	(0.01)		
<u><i>Men</i></u>			<u><i>Women</i></u>	
<i>Age</i>				
<i>Average</i>	35.2	29.5	35.4	28.3
	(0.03)	(0.17)	(0.03)	(0.39)
<i>Median</i>	34	27	34	24
<i>Schooling years</i>				
<i>Average</i>	8.5	7.2	7.9	8.4
	(0.01)	(0.06)	(0.01)	(0.15)
<i>Median</i>	9	6	9	9
<i>Live in Rural Area</i>	22%	45%	22%	34%
	(0.00)	(0.01)	(0.00)	(0.02)
<i>Agricultural workers</i>	20%	42%	5%	7%
	(0.00)	(0.01)	(0.00)	(0.01)
<i>Labor force participation</i>	87%	89%	42%	39%
	(0.00)	(0.01)	(0.00)	(0.02)
<i>Wage Earners</i>	68%	60%	28%	24%
	(0.00)	(0.01)	(0.00)	(0.01)
<i>Unemployment rate</i>	1.8%	3.9%	1.0%	2.0%
	(0.00)	(0.00)	(0.00)	(0.00)
<i>Hourly wage in 2006 dollars</i>				
<i>Average</i>	1.95	1.38	1.88	1.44
	(0.00)	(0.02)	(0.01)	(0.08)
<i>Median</i>	1.43	1.15	1.32	1.08
<i>Hourly wage relative to the quarter average</i>				
<i>Average</i>	1.01	0.72	0.97	0.75
	(0.00)	(0.01)	(0.00)	(0.04)
<i>Median</i>	0.74	0.60	0.68	0.56

Table 1: ENET 2000-2004 Summary Statistics

Source: ENET. Individuals aged 16 to 65 years old. Standard errors in parentheses. 2000 only includes the last three quarters and 2004 only the first three quarters. The construction of wages follows the lines of Chiquiar and Hanson (2005). The ENET asks Mexicans for their wage in the week previous to that in which the survey is performed or, if the individual did not work that particular week, for the usual wage. The figure is then brought to the monthly level. In order to prevent wages to refer to different time periods, the observations for individuals who reported usual rather than actual wage income are dropped. I follow Chiquiar and Hanson (2005) in dropping observations of individuals who worked more than 84 hours or less than 20 hours per week and then the highest and lowest 0.5 percent of observations to eliminate outliers. Finally, the observations for people who worked in the United States (mostly border workers) are also dropped. Real wages are constructed with inflation data from the INPC series, Mexican CPI, in Banxico (www.banxico.org.mx), the Mexican central bank. These are quarterly averages based on June 2002 and brought to December 2005 with an index of 116.301. The exchange rate, from the International Financial Statistics of the IMF, corresponds to the 1 January 2006 and it is 10.7777 pesos per dollar. Following Chiquiar and Hanson (2005), hourly wages are computed by dividing the monthly wage income reported in the ENET by 4.5 times the number of hours worked in the previous week. The quarter average is computed pooling observations for men and women and migrants and non-migrants. Individuals are considered to live in a rural area when their locality has less than 2,500 inhabitants according to the 2000 Mexican Census.

The definition of emigrants as those who emigrate a quarter later turns out not to be relevant. Defining the emigrant group as those who emigrate two, three or four quarters after their first interview does not alter the results. In fact, the hypothesis of an Ashenfelter dip in migration, that is, that there is a sudden drop in wages before the migration decision takes place is rejected for the yearly horizon that the dataset allows to explore: there are no changes in the wages of future emigrants during the four quarters anteceding the decision to leave for the United States.

Table 1 also shows other important characteristics of the emigrant population with respect to non-migrants. First, the proportion of men in the emigrant sample reaches 81 per cent whereas they represent only 47 per cent of the non-migrant population. Emigrant men are significantly younger (29.5 versus 35.2 years old) and less educated than non-migrants on average. The median education of emigrants is 6 years of schooling, which corresponds to primary school completion. For non-migrant men, the median education is 9 years of schooling, which corresponds to finished middle school, that is, “*Secundaria*” in Spanish, which is often mistaken by high school completed in US sources (Ibarrarán and Lubotsky (2007)). The extent to which these observable characteristics explain the negative selection result will be assessed in the next section.

Another striking difference in table 1 is the high percentage of emigrants coming from rural areas relative to non-migrants (45 versus 22 per cent for men). Rural areas are defined by localities with less than 2,500 inhabitants in the 2000 Mexican Census. The high proportion of agricultural workers in the sample is consistent with this higher prevalence of emigration in rural Mexico with respect to urban areas. Despite this observation, it must be noted that the majority of emigrants actually comes from urban Mexico so that it could be misleading, as it will be shown in the next section, to extend results found in rural Mexico (i.e. results based on the Mexican Migration Project) to the whole country.

Finally, there are no important differences in labor force participation between migrants and non-migrants. The fact that the percentage of wage earners is higher in the case of non-migrants is explained by the over-representation of rural Mexico migrants in the emigrant sample. The percentage of wage earners in rural Mexico is significantly lower both for emigrants and non-migrants. A similar consideration applies to the differences in the unemployment rate, which is shown to be higher for migrants (3.9 per cent) than for non-migrants (1.8 per cent).

The second part of table 1 shows the same summary statistics for non-migrant and migrant women. The main difference with men is found in the education level. Whereas emigrant men are less educated than non-migrant men, the reverse is true for emigrant women, who are slightly more educated on average (8.4 versus 7.9 years of schooling) than non-migrant women although the two groups share the same median education level of 9 years of schooling. In general, the main inconvenience for the study of emigrant selection among women is the low percentage of wage earners so that it can be more relevant to look at schooling levels rather than at wages. The low levels of labor force participation among women suggest that many of them could be considered as tied-movers, that is, as individuals traveling to the United States to join their spouses, so that economic considerations would not weigh in their decision as much as for men. On the other hand, it is also evident in table 1 that migrant women are younger than migrant men so that another reason for the positive selection in schooling levels could be the fact that younger generations of Mexican women are getting more educated.

More details about how these magnitudes in the ENET compare to other traditional data sources can be found in the appendix.

A potential problem is that the ENET does not record those individuals whose complete household emigrates to the United States but only those for which one of the household members still remains in Mexico in a second or further round of interviews. Individuals migrating with their whole household might be expected to have higher education and wages than those who migrate leaving their families behind and this would create a bias towards finding negative selection. McKenzie and Rapoport (2007b) report that 14.4 per cent of male migrants from 18 to 45 years old arrived in the last two years in the 2000 US census had their spouse present. This number is 48 per cent for women. However these figures might be overestimating the true bias in the number of ENET migrants as long as spouses did not migrate exactly at the same time or underestimating it in the case of single-person households (2.26 per cent of total households in the ENET). Ibararán and Lubotsky (2007) compare the US and Mexico 2000 census and conclude that full households of Mexicans in the US account for between 16 and 25 per cent of the total population of Mexicans in the US. Again, this does not necessarily mean that all members of the household left at the same time. US sources (see section 5.1 for definitions) show that Mexican men in the US with their spouse present tend to have 0.4 more years of education than those without (although the opposite

result is obtained for women) so that the bias is clearly insufficient to overturn the negative selection result in terms of schooling years (there is a difference of 1.3 years of education in table 1). Another Mexican survey: the Encuesta sobre Migración en la Frontera Norte (EMIF) or Northern Frontier Migration Survey shows that only 8.6 per cent of individuals crossing the border by land in the period 2000-2003 with the purpose of emigrating carried all of their household members with them.¹⁷ The EMIF is the only nationally representative source that allows a comparison of individuals who travel with the whole household with those who travel with only part of it. The average and median schooling and age of these two groups are not statistically different (see the data appendix for a more detailed comparison of the ENET and the EMIF). However, the point estimates suggest that the bias goes in the expected direction. The average education years for men traveling with their whole household are 8.2, contrasting with 6.8 for the rest of men. For women, the results go again in the opposite direction, women traveling with their whole household have on average 6.1 years of education whereas those who leave household members behind reach 7.3 years. Although the differences in point estimates are higher in the EMIF, the total effect of this bias in the general sample is still insufficient to overturn the negative selection result. You would need to assume an undercount rate of 25 per cent (the highest in the range of Ibararán and Lubotsky (2007)) and 10 years of average education years (way above the estimates provided by the EMIF or even US sources) to be able to cut the negative selection result in terms of education years in half. The only concern may appear for women since in their case it is 30 per cent of emigrants who travel with their whole families, contrasting with only 6 per cent of men. Thus, the results about women in this paper should be taken with more caution than those for men.

¹⁷The EMIF interviews people at traditional border crossing points representing 94 per cent of total crossers, according to estimates from CONAPO (Consejo Nacional de Población or National Population Council), the Mexican agency that runs the survey (CONAPO (2006)). How representative can this be from total Mexican emigration? The Módulo sobre Migración de la Encuesta Nacional de Empleo (ENE) or Migration Module of the National Employment Survey in Mexico in 2002 estimated that around eighty per cent of Mexican migrants crossed the border by land (INEGI (2004)) so that the EMIF would be able to cover around three quarters of total Mexican emigration per year. The number in the text is calculated over 7,799 observations of border crossers who declared they wanted to work or stay longer than a year in the United States.

4 Selection of Emigrants from Mexico to the United States

This section presents new evidence on emigrant selection, namely the fact that individuals emigrating from Mexico to the United States during the 2000-2004 period were in general less productive, as reflected in their wages, than individuals who chose not to emigrate.

The main result of this paper comes from the direct comparison of the wage distributions of migrants and non-migrants. The great advantage of the ENET for the study of selection is that the wage of migrants in the quarter previous to that in which they emigrated can be observed at the same time as non-migrants' wages. Table 1 already suggests that negative selection in terms of wage levels prevails on average. However, average wages could be hiding possibly interesting differences in the distribution. Figure 3 shows that this is not the case, by jointly graphing the wage distribution for emigrants and non-migrants for men and women. Figure 3 groups the data from the whole period by dividing actual real wages by the quarter average (so as to avoid time trend effects) and then taking the log distribution of these relative wages.

The graphs represent the kernel density estimate¹⁸ of the distribution of the logarithm of real hourly wages¹⁹ relative to its quarter average registered for the group of migrant and non-migrant men and women aged 16 to 65 years old in the period going from the second quarter of 2000 to the third quarter of 2004.

Concentrating first on men, figure 3 confirms the existence of negative selection in the studied period. The distribution of wages of future migrants lies clearly to the left of the distribution of wages for non-migrants. If wages represent productive characteristics of individuals in Mexico, this can be seen as evidence that emigrants tend to be less productive than non-migrants.

The corresponding wage distribution for women in figure 3 suggests that there was also negative selection in terms of wage levels for female emigrants although, as we saw in table

¹⁸The estimated density is $\hat{g}(w) = \frac{1}{hN} \sum_{i=1}^N K\left(\frac{w-w_i}{h}\right)$ where N is the number of observations. $K(u) = \frac{3}{4}(1-u^2)$ for $-1 < u < 1$ and $K(u) = 0$ otherwise is the Epanechnikov kernel, where $u = \frac{w-w_i}{h}$. The optimal bandwidth (Silverman (1986)) is $h = 0.9\hat{\sigma}N^{-\frac{1}{5}}$ with $\hat{\sigma} = \min\left\{S, \frac{IQR}{1.349}\right\}$ where S is the sample standard deviation and IQR is the inter-quartile range. To prevent over-smoothing and following Leibbrandt, Levinsohn, and McCrary (2005), I use a bandwidth which is 0.75 times this optimal level.

¹⁹Using monthly wages does not alter the results.

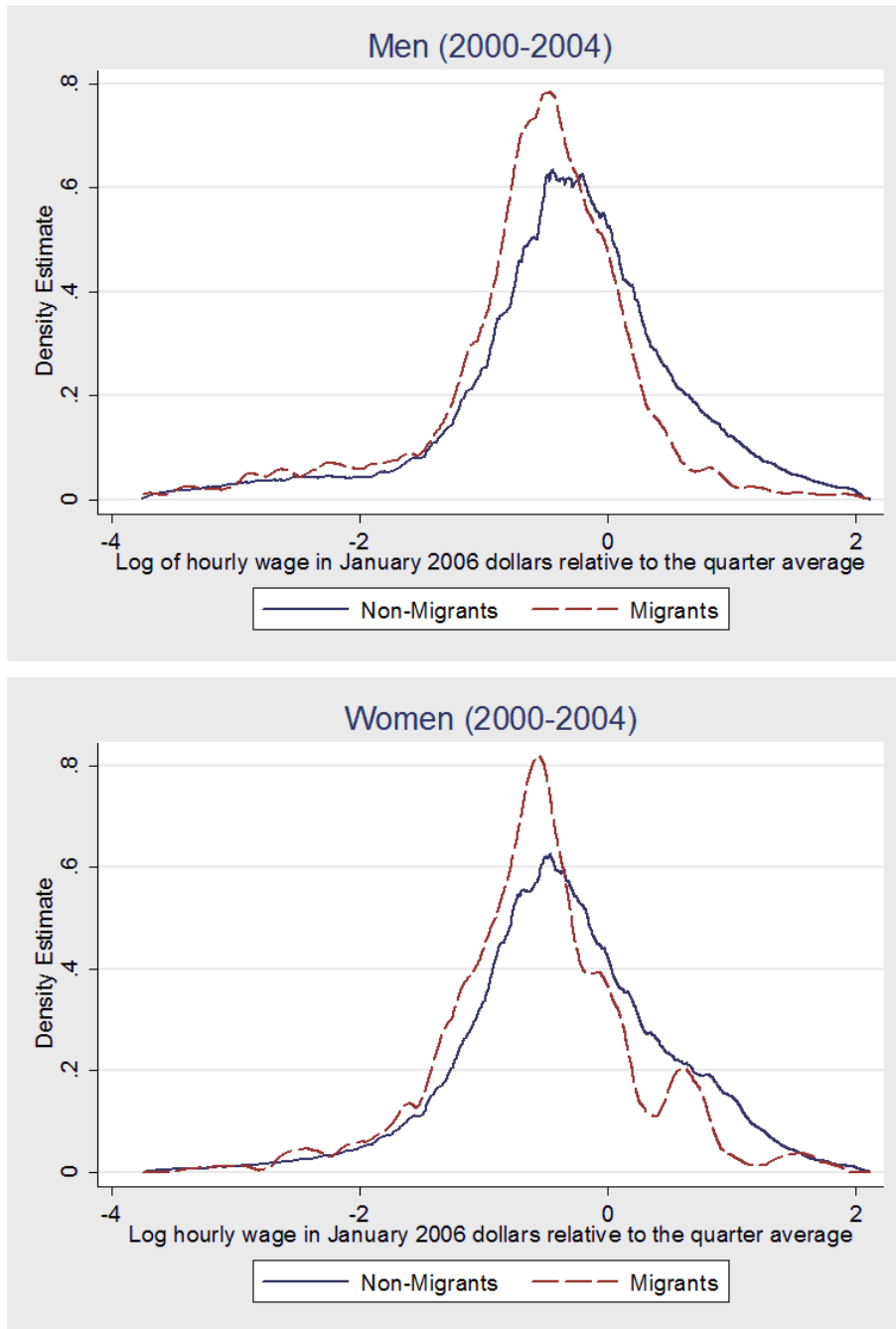


Figure 3: Wage distribution of migrants and non-migrants

Source: ENET. Log of the hourly wage relative to the quarter average. See table 1 for the construction of wages. For the estimation of the kernel densities, I use an Epanechnikov kernel instead of the Gaussian kernel preferred by Chiquiar and Hanson (2005) because the former is the most efficient in approximating the true density (Silverman (1986)). To prevent over-smoothing, I follow Leibbrandt, Levinsohn, and McCrary (2005) in using a bandwidth which is .75 times the optimal.

1, the number of observations is much lower.

To provide a clearer picture of the selectivity in Mexican emigration, the difference of the two densities (migrant minus non-migrant) plotted in each graph in figure 3 is plotted alone in figure 4.

The vertical solid black line represents the median wage for each category. Positive mass of the density difference to the left of the median means negative selection predominates whereas positive mass to the right would have been an indication of positive selection. Focusing on men, figure 4 confirms what was explained when dealing with figure 3. Negative selection clearly characterizes the emigrant wage distribution. For women, negative selection is also confirmed.

A first possible concern with these graphical results is whether they are statistically significant. One way of testing whether the wage distributions for emigrants and non-migrants come from the same underlying wage distribution is to perform a Kolmogorov-Smirnov test for the equality of distribution functions. The null hypothesis is precisely that both distribution functions are equal. The way the test operates is by calculating the differences between the empirical distribution functions, with no need for additional assumptions except for that of continuity. The empirical distribution functions are represented in figure 5.

Figure 5 is another way of presenting the same information reported in figures 3 and 4 as an empirical distribution function rather than a density function, with the advantage that no kernel density estimation is involved. Negative selection appears again clearly for both men and women. The D statistic from the Kolmogorov-Smirnov test based on these empirical distribution functions is 0.1566 for men and 0.0710 for women, both rejecting the hypothesis that the wage distributions for migrants and non-migrants are the same at a 0.1 per cent significance level.

In order to add the time dimension and for completeness, it is interesting to compute the degree of selection as defined in section 2, that is, the average log wage obtained by future emigrants minus the average log wage obtained by non-emigrants in a given quarter. The results can be observed in figure 6.

Figure 6 presents average log hourly wages for Mexican men aged 16 to 65 years old. They are classified in two groups: migrants are those who are reported to have left for the United States a quarter after the wage observation takes place whereas non-migrants are those who remain in Mexico. The average log hourly wage increases from 0.23 at the beginning of the

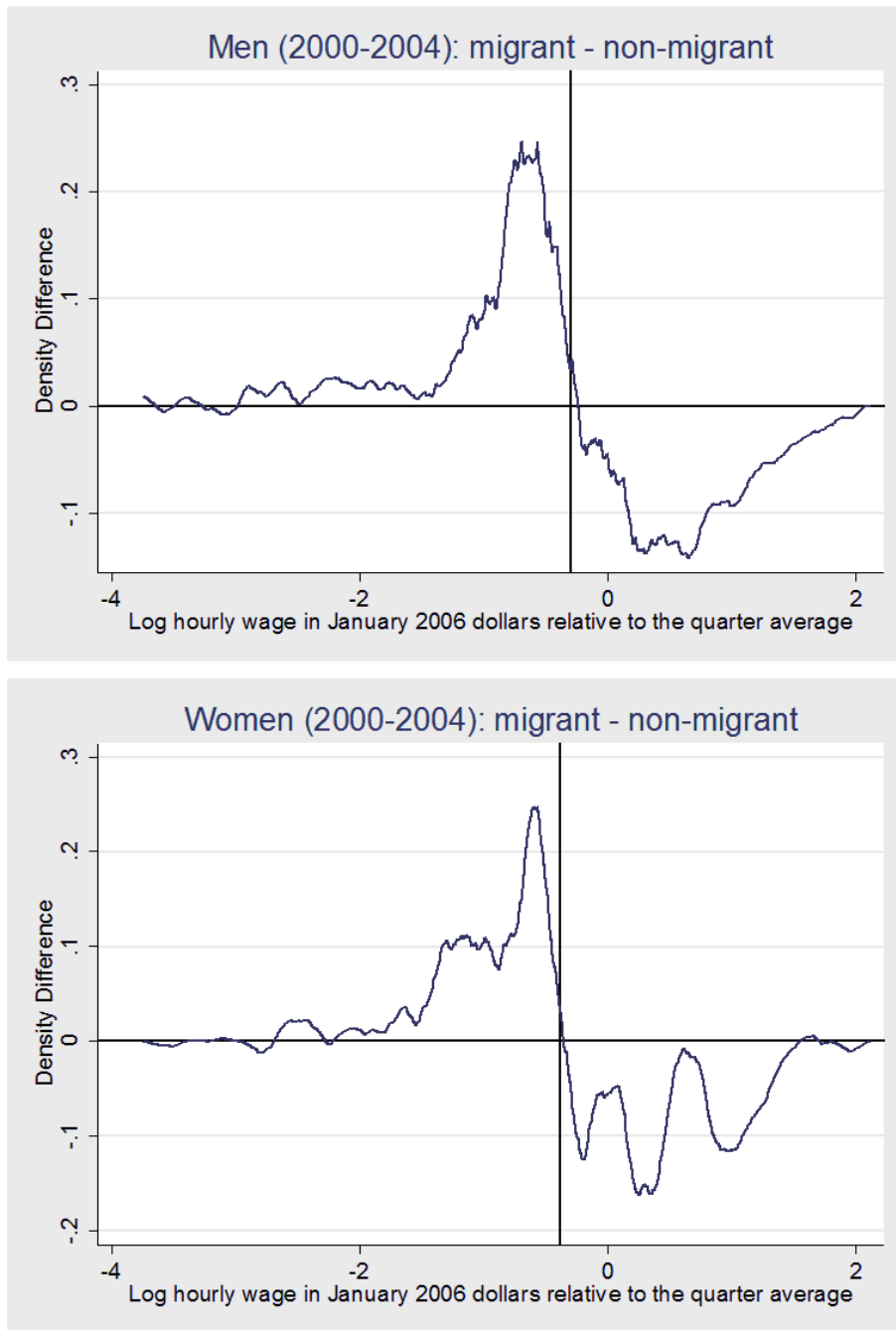


Figure 4: Wage distribution of migrants minus wage distribution of non-migrants
 Source: ENET. Migrant minus non-migrant wage densities computed in figure 3. See figure 3 for an explanation. The solid black vertical line represents the median of the log of the relative wage distribution.

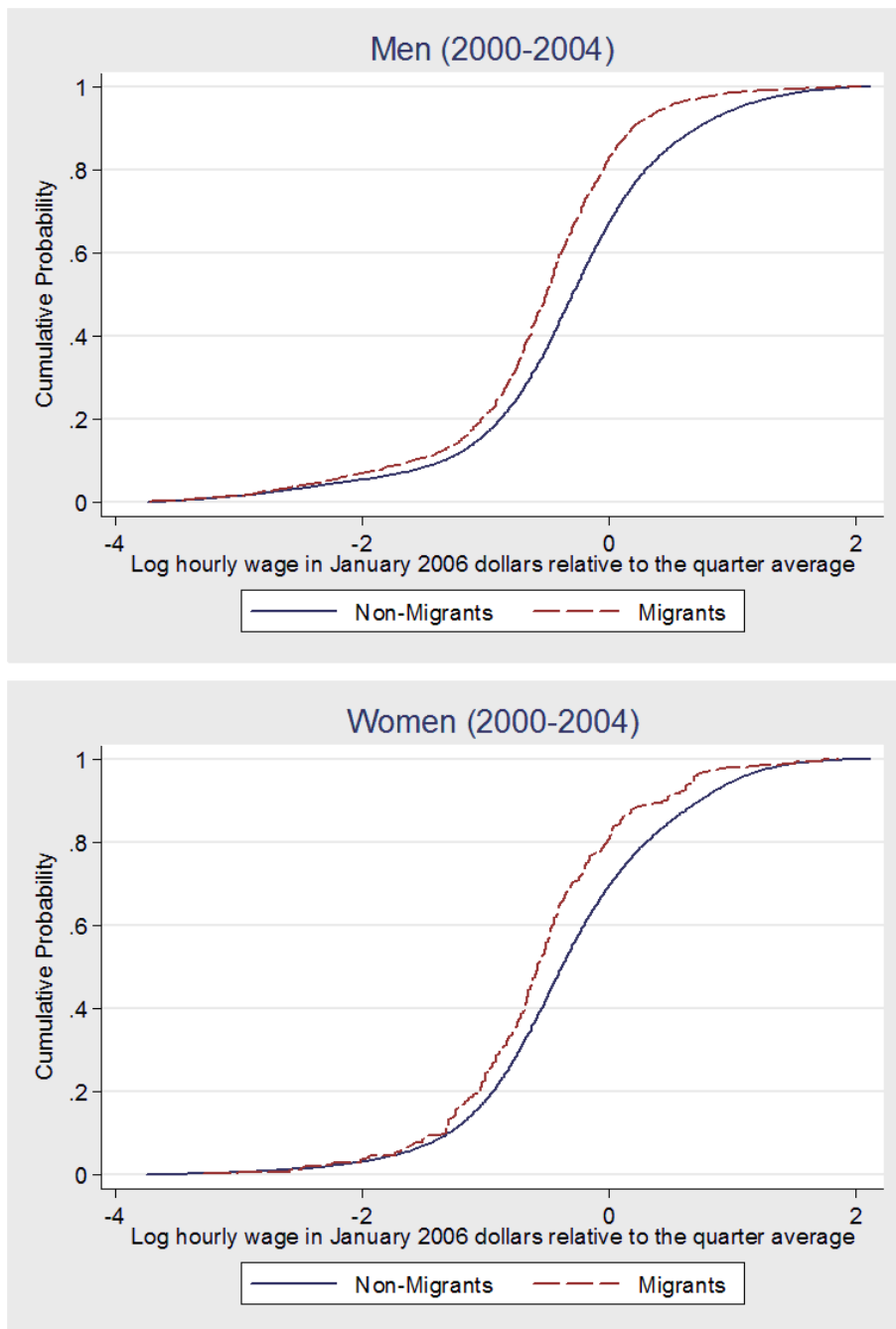


Figure 5: Distribution functions of migrant and non-migrant wages

Source: ENET. Empirical distribution functions of the log of the hourly wage relative to the quarter average. See table 1 for the construction of wages.

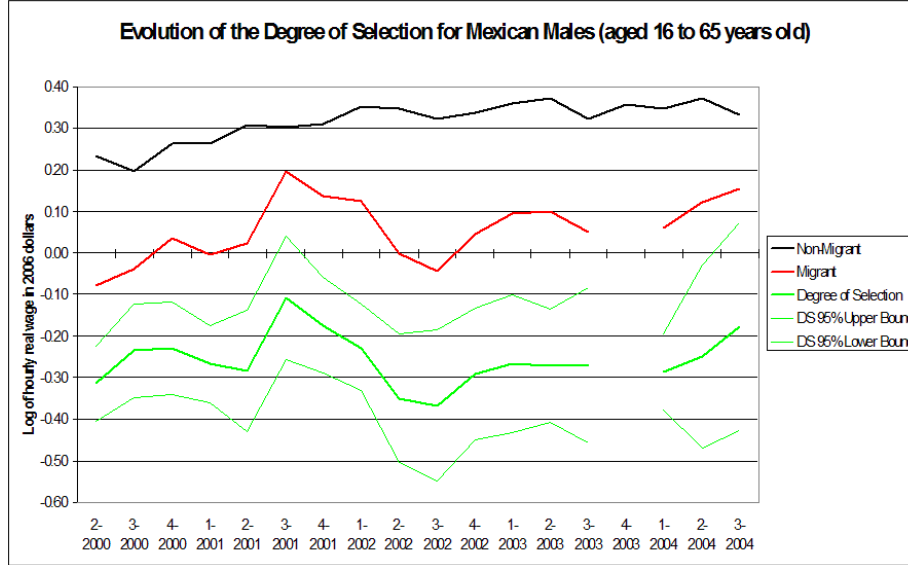


Figure 6: Distribution functions of migrant and non-migrant wages

Source: ENET. Log of hourly wages. See table 1 for the construction of wages. The fourth quarter of 2003 is not reported due to a dataset problem (see footnote 20 in the text). The degree of selection is the difference between the migrant and the non-migrant average log wage.

period to 0.33 at the end for non-migrants, with a low of 0.20 in the third quarter of 2000 and a high of 0.37 in the second quarter of 2004. The corresponding wage for migrants is regularly below that of non-migrants and significantly so for 15 of the 18 quarters studied at a 95 per cent confidence level. Only the last quarter of 2003 shows a higher wage for migrants than for non-migrants²⁰. Abstracting from this exception, the degree of negative selection in terms of the average log wage oscillates during the period between a low of 0.11 in the third quarter of 2001 and a high of 0.37 in the third quarter of 2002. The variation across periods is not significant so that grouping the data as it was done in the previous figures is justified. However, for robustness purposes, the analysis in figures 3 to 5 was replicated at a yearly and quarterly level, with analogous results.

The wage an individual obtains, in addition to market conditions, responds to her observable and unobservable characteristics. In this sense, it is interesting to look at how some basic observable characteristics, such as age or experience and schooling levels, relate to the

²⁰The fourth quarter of 2003 does not register permanent emigrants but only temporary ones (see definitions in the appendix). The results of the paper are robust to including or excluding this quarter.

emigration decision. In other words, how do emigrants select in terms of observable characteristics? This exercise is also useful to check whether the selection that takes place outside the labor market (among individuals who do not report earning a wage) is different from what was found above. Thus, the analysis that follows refers to all individuals aged 16 to 65 years old in the ENET sample and not just to wage earners (the results are analogous if the sample is restricted to wage earners).

Figure 7 shows the age distribution of migrant and non-migrant men and women. In both cases, it appears clear that emigrants tend to be younger than the rest of the population, as it was already shown by the calculation of the average and the median in table 1. The male distribution concentrates the higher proportion of emigrants from 18 to 20 years of age, with the wider group 16 to 20 constituting more than a quarter of total male emigrants to the United States. This is even more accentuated for women, where the emigrant mode is also at 20 and the wider group from 16 to 20 years includes more than 30 per cent of the female emigrant population in the period. Expanding the age window from 16 to 25 years accounts for 45 per cent of male emigrants and more than 55 per cent of female emigrants. In fact, emigrant males outnumber (relatively) non-migrants in almost all ages from 17 to 39. In the case of females, the range of higher concentration of emigrants goes from 16 to 28. This confirms the fact that migration is a long term investment which is more profitable for younger individuals.

This concentration of emigrants at younger ages could raise the concern that the negative selection result just reflects seniority so that selection could be positive within age groups. In fact, this is not the case. In calculations available from the author upon request, it can be shown that selection is negative for all age groups although it appears that the degree of negative selection is less pronounced for younger categories (the difference in the degree of selection among different age groups is not statistically significant nonetheless).

The education level is the other main observable characteristic and it is sometimes identified as a proxy for the x used in the theoretical model above summarizing productive characteristics (Ibarrarán and Lubotsky (2007)). The selection of emigrants in terms of educational attainment is presented in figure 8.

Male Mexican emigrants are relatively more abundant than non-migrants in all education levels from 3 to 9 years of schooling. In addition, the mode for emigrants is at 6 years of schooling (completed primary school) whereas the mode for non-migrants is situated at 9

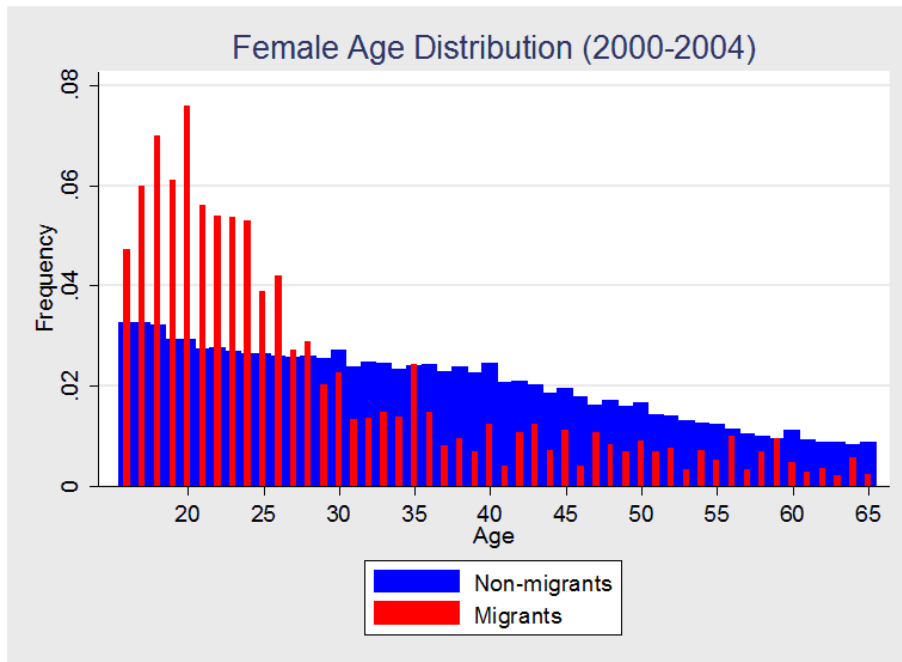
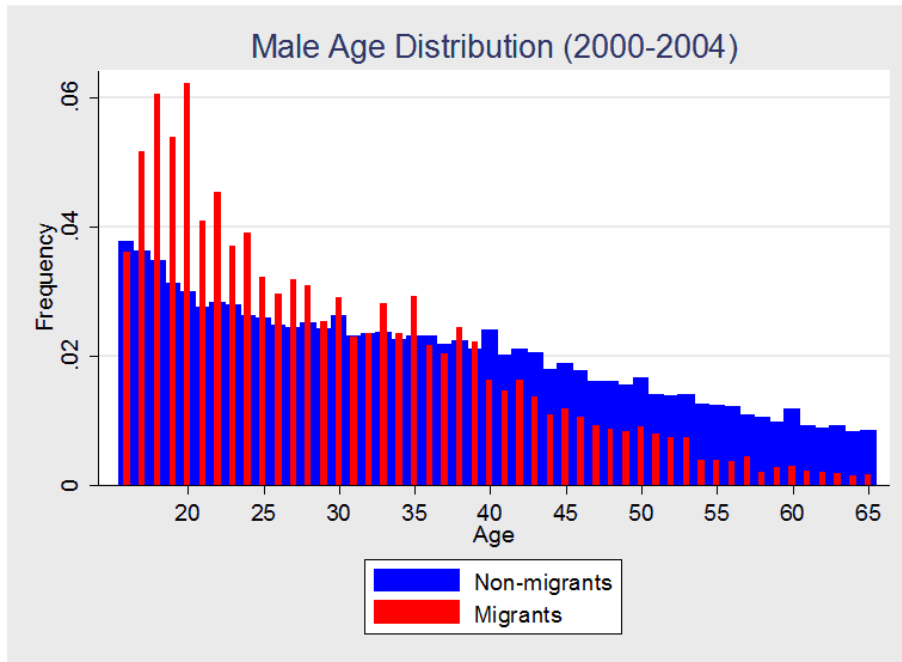


Figure 7: Age distributions of migrants and non-migrants
 Source: ENET. The histograms represent the percentage distribution of migrants and non-migrants at different ages.

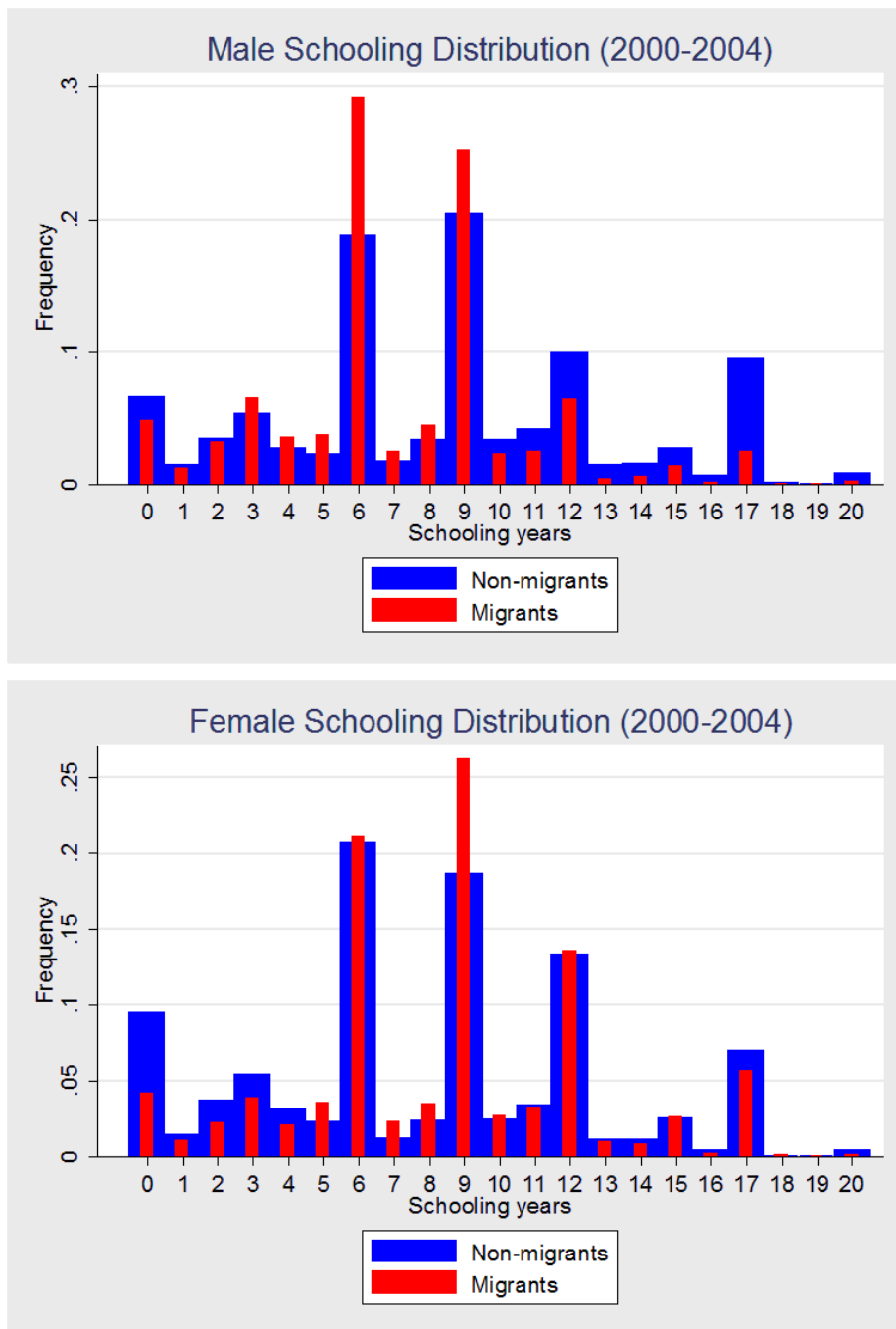


Figure 8: Schooling distributions of migrants and non-migrants

Source: ENET. The histograms represent the percentage distribution of schooling years for migrants and non-migrants. The last bin includes individuals with 20 or more years of schooling.

years of schooling (completed middle school). It must be pointed out that both would be classified as high school dropouts according to another usual division in the literature (Ottaviano and Peri (2006)). In this sense, the education groups of high school graduates and beyond (12 and more years of schooling) present a higher percentage of non-migrants than of migrants. Despite this, the lowest education groups (no schooling and primary school dropouts with up to 2 years of schooling) present lower concentration of migrants than of non-migrants. Still, the general conclusion from these graphs is the existence of intermediate to negative selection of male emigrants in terms of schooling years.

Contrary to other cases, the female schooling distribution of migrants and non-migrants turns out to be different from the male one. Female emigrants relatively outnumber non-migrants in most categories from 5 to 15 years of schooling. In addition, and in reverse to the male case, the female emigrant mode is situated at 9 years of schooling whereas the non-migrant mode remains at 6 years of schooling despite a slow increase in female schooling registered during the period.

To sum up, the ENET data show clear evidence of negative selection in terms of wages for men and women, intermediate to negative selection in terms of education for men and slightly positive selection in terms of education for women. The following step is to explain why previous literature obtained different results and whether those can be reconciled with the findings above.

5 Differences with Previous Literature

In contrast to the negative selection result presented in the previous section, recent papers using two different sources recently found that positive selection characterized migration flows between Mexico and the United States. On the one hand, Chiquiar and Hanson (2005) and Mishra (2007) combined the US and Mexican Census. On the other hand, McKenzie and Rapoport (2007a) and Orrenius and Zavodny (2005) used data from the Mexican Migration Project (MMP). In addition, Chiquiar and Hanson (2005) tried to address some of the possible problems in their source by complementing it with the MMP.

The main problem of using the US Census or, more generally, US sources to study selection is the known fact that the Census under-counts immigrants, especially those who are undocumented and possibly low skilled (Hanson (2006)). As for the MMP, its principal

inconvenience is the fact that it is not nationally representative so that results based on it might not be generalizable to Mexico as a whole, as we will see below.

The ENET is nationally representative and homogeneous in the sense that the same source is used to gather information on both migrants and non-migrants. Only those emigrants who do not leave anybody behind (all of the household leaves) are missed by the ENET (see section 3 for a discussion). Apart from this deficiency, the ENET is an “ideal” dataset to study emigrant selection since the questionnaire is performed right before the emigrant decides to leave. However, the ENET is not good at providing information about what the emigrant will do later or, more exactly, on whether the emigrant from Mexico will become a permanent immigrant in the United States. This can be a source of bias in the comparison with studies based on the characteristics of migrant stocks since the ENET flow includes both return migrants and temporary migrants who go back and forth often between Mexico and the United States.

In the case of return migrants, Borjas and Bratsberg (1996) and Lacuesta (2006) show that their characteristics tend to be somewhere in between those of migrants and non-migrants. Using the US and Mexican Census, Lacuesta (2006) compares returned Mexican migrants with non-migrants and with those who remain in the United States.²¹ He concludes that return migrants tend to be negatively selected with respect to Mexican immigrants in the US but positively selected with respect to non-migrants who remained in Mexico. However, return migrants in Lacuesta (2006) would also be very positively selected with respect to the emigration flow recorded by the ENET since they earn higher wages and have more education years than non-migrants and thus much higher wages and much more education years than ENET migrants. The ENET also provides information on return migrants, but only when they come back to the same household where they previously lived. Subject to this bias, sampled return migrants in the ENET amount to 32.7 per cent of total emigrants (both in terms of total population and in terms of working age males). Their average education level is 6.9 years, below the 7.2 registered for emigrants. This would imply that the average education level of “net” migrants, that is, the average level of emigrants corrected by return migration, would be 7.4 years, still well below the 8.5 years registered for non-migrants.

In the case of temporary migrants, who come back to Mexico in less than a year, Hanson

²¹Reyes, Johnson, and Van Swearingen (2002) report a return migration rate of 11 per cent after one year from the 2000 Mexican Census.

(2006) pointed out that studies concentrating on temporary or seasonal migrants would tend to find results biased towards negative selection since this type of flows would be characterized by lower education and a higher percentage of men than more permanent migration. This observation came from the study of the MMP, where the profile of permanent migrants differs from that of temporary migrants in this direction (Reyes (1997)). Unfortunately, the information on temporary versus permanent migrants in the ENET is not reliable (see appendix) since it is based on “ad-hoc” definitions.²² Nevertheless, if the result was only due to the presence of temporary and seasonal migrants, the degree of negative selection should be more pronounced in the periods of the year when seasonal migration takes place. This is not the case and measuring the degree of selection by quarter does not lead to significantly different results, that is, the degree of selection does not show any seasonal pattern, whether the data are grouped by quarter²³ or, as in figure 6, disaggregated data are used. This observation is consistent with the finding that seasonality is ceasing to characterize migration flows from Mexico to the United States in the last years (Marcelli and Cornelius (2001)). For example, Angelucci (2005) points out that one of the effects of tougher border enforcement by the US government was to stop the circular flow of migrants. At the same time that tougher enforcement contributed somehow to reduce entry flows, it also caused a reduction in exit flows, making migration more permanent.

5.1 Comparison with Chiquiar and Hanson (2005)

When comparing the ENET data to sources based on the Mexican and US census, the main difference that must be taken into account is that between a flow and a stock variable. The ENET measures the emigration flow whereas the US census reflects the immigration stock. In a study of selection, it can be argued that it is more appropriate to look at the flow rather than the stock since, for example, Mexican immigrants may acquire new education or other skills during the time they are in the United States. This is why Chiquiar and Hanson (2005) produce robustness checks for their estimates by reducing their population of interest to those who have not been long in the United States. Still, despite the different

²²However, estimating the degree of selection for permanent and temporary migrants as defined in the ENET produces results that actually contradict Hanson (2006) observation. Negative selection is clear for permanent migrants whereas it is not so obvious for temporary ones.

²³Results available from the author upon request.

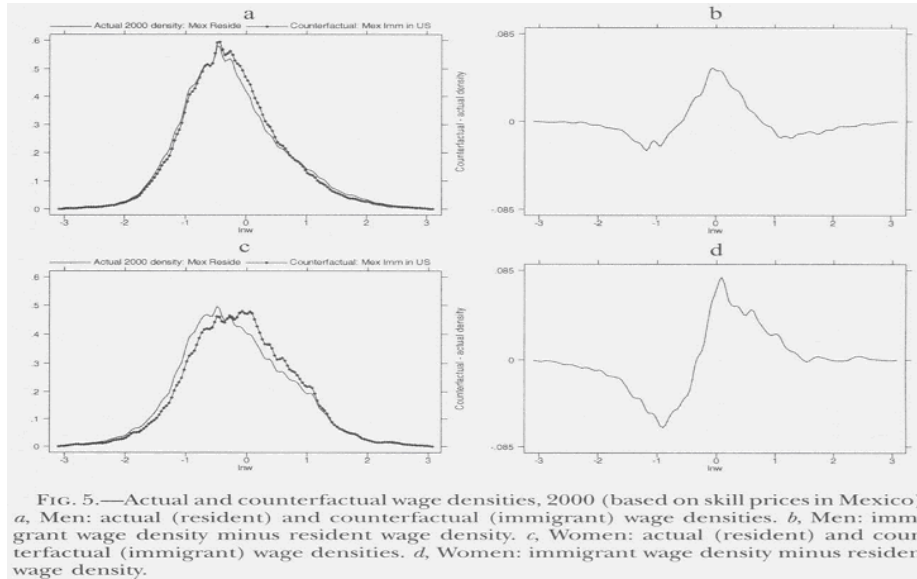


Figure 9: Chiquiar and Hanson (2005) main result

Source: Chiquiar and Hanson (2005). Data on Mexican residents from the 2000 Mexico Census. Data on Mexican immigrants from the 2000 US Census.

concepts (stock versus flow), it remains surprising that Chiquiar and Hanson (2005) result of positive selection is opposite to that of the ENET. We now turn to studying the source of this difference.

Chiquiar and Hanson (2005) adopt DiNardo, Fortin, and Lemieux (1996) methodology to build a counterfactual wage density reflecting how much Mexican immigrants would earn if they were to go back to Mexico according to their skill level recorded in the US Census in 1990 and 2000. They find that this counterfactual wage density lies slightly to the right of the actual Mexican wage distribution, which they interpret as a proof of positive selection. The advantage of the ENET is that the wage of Mexican emigrants can be observed right before they move to the United States so that there is no need to build a counterfactual wage density. One just needs to compare the wage density of non-migrant Mexicans in a given period with the actual wage density of future migrants. For comparison purposes, Chiquiar and Hanson (2005) main result is reproduced here as figure 9.

Panel a in figure 9 compares the logarithmic wage (hourly wages) distribution of Mexican males living in Mexico at the time of the 2000 Mexican census (solid line) with the counterfactual wage distribution that Mexicans living in the US in 2000 according to the US

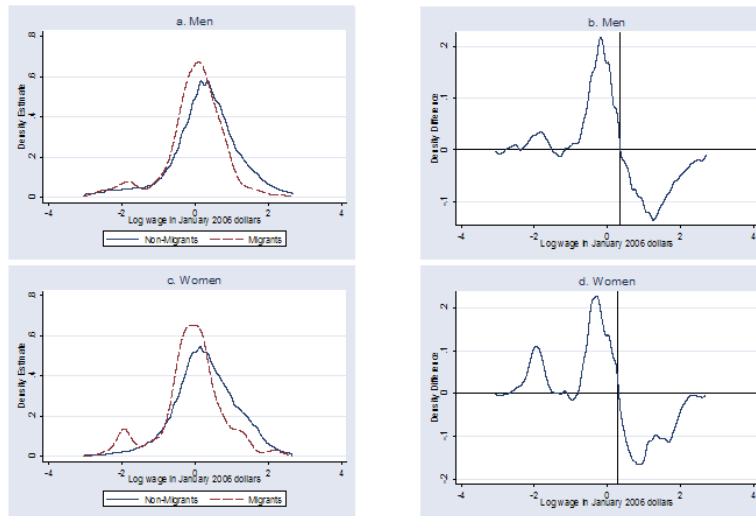


Figure 10: 2000 Migrant and non-migrant wage distributions (a,c). Migrant wage density minus non-migrant wage density (b,d)

Source: ENET. Wages are computed as in table 1 but the sample is reduced to observations from the year 2000 of individuals aged 21 to 65 years old and an optimal Gaussian kernel with optimal bandwidth is used instead of the Epanechnikov to adjust more to Chiquiar and Hanson (2005) methodology.

census would generate in case they came back to Mexico without affecting the wage distribution. Since the counterfactual wage distribution lies slightly to the right of the actual wage distribution, Chiquiar and Hanson (2005) conclude that there is positive selection of male immigrants from Mexico to the United States. Panel b shows how the density difference is slightly negative for low wage levels and positive for medium-high wage levels. Panels c and d prove that the positive selection result is stronger for women.

The data come from the US and Mexico 2000 censuses, which both reflect the situation in the first half of 2000 whereas the ENET dataset only starts in the second quarter of 2000 with information on emigrants who left in the third quarter of the same year. As a result, no direct comparison is possible between Chiquiar and Hanson (2005) and these data. In any case, using only 2000 data from the ENET and following Chiquiar and Hanson (2005) methodology, the selection picture emerging from the ENET can be seen in figure 10.

Wage data are generated in the same way as those presented in figures 3 and 4 without dividing by the quarter average except for the fact that age is restricted to be from 21 to 65 years old as in Chiquiar and Hanson (2005). Optimal bandwidth with a Gaussian kernel (Silverman (1986)) is used instead of 0.07 log wage points to allow for different sample sizes of

migrant and non-migrant observations²⁴. Confirming what was found in figures 3 and 4 with a different kernel (see footnote 18) and for a different time period, there is negative selection of both male and female Mexican emigrants in terms of the wage they earned before they decided to migrate, compared with the wage earned by those who decided not to migrate.

The reasons why the negative selection result in this paper differs from Chiquiar and Hanson (2005) positive selection result can be more carefully analyzed. First, it is possible that unobservable components matter more in the migration decision than observables, as Borjas (1999) suggests, so that the problem is methodological because DiNardo, Fortin, and Lemieux (1996) non-parametric estimation technique does not reflect the effect of unobservables. A second hypothesis is simply about the quality of the data. The methodology can be correct but the combination of data from the US and Mexico Census biases the results either because undocumented migrants are not captured by the US Census or because Mexican immigrants tend to over-report their education level in the US Census relative to the education categories in the Mexican Census, as it is suggested by Ibararán and Lubotsky (2007).

5.1.1 Selection on Observables. Disregarding the Direct Information on Wages from the ENET

In order to address the first hypothesis, I assume that the ENET does not provide data on the wages of emigrants who leave for the United States. Concentrating in the case of males, the actual wage distribution of non-migrants computed in figure 3 can be compared now not to the actual wage distribution of emigrants but to a counterfactual wage distribution built from their observable characteristics, following DiNardo, Fortin, and Lemieux (1996). To fix ideas, denote by $g_{M,0}(w)$ and $g_{NM,0}(w)$ the relative wage distribution for migrants (M) and non-migrants (NM) observed at location 0 (w will represent the logarithm of the relative wage in this section). The analysis in the previous section was based in the direct estimation and comparison of these two densities (figures 3 and 4). They can also be rewritten as:

$$g_{i,0}(w) = \int f_{i,0}(w|x) h_{i,0}(x) dx; i = NM, M$$

²⁴The resulting optimal bandwidths are: 0.07 for non-migrant males (sample size of 282,706); 0.15 for migrant males (1,120 observations); 0.07 for non-migrant females (138,065 observations); and 0.20 for migrant females (147 observations).

where $f_{i,0}(w|x)$ represents how the wage responds to changes in characteristics x (it refers only to observable characteristics) and $h_{i,0}(x)$ is the density of characteristics at location 0 for individuals in situation i . Now, instead of directly observing $g_{M,0}(w)$, assume that this has to be estimated from the observable characteristics of emigrants. Formally, the required counterfactual is:

$$\hat{g}_{M,0}(w) \equiv \int f_{NM,0}(w|x) h_{M,0}(x) dx$$

that is, the estimated wage distribution of emigrants will be based on the way observable characteristics of non-migrants are rewarded: $f_{NM,0}(w|x)$. In order to do this, DiNardo, Fortin, and Lemieux (1996) suggest the following. First, rewrite the density as:

$$\hat{g}_{M,0}(w) = \int f_{NM,0}(w|x) h_{M,0}(x) dx = \int f_{NM,0}(w|x) h_{NM,0}(x) \frac{h_{M,0}(x)}{h_{NM,0}(x)} dx$$

This is equivalent to reweighting the non-migrant wage distribution by the factor $\theta \equiv \frac{h_{M,0}(x)}{h_{NM,0}(x)}$, which can be computed using Bayes' theorem as:

$$\theta = \frac{h_{M,0}(x)}{h_{NM,0}(x)} = \frac{\frac{P(M|x)}{1-P(M|x)}}{\frac{P(M)}{1-P(M)}}$$

Following Chiquiar and Hanson (2005), $P(M|x)$ can be estimated from a logit model of the probability of emigration²⁵ regressed on observable characteristics whereas $P(M)$ refers to the proportion of emigrants in the sample. The result from estimating $\hat{g}_{M,0}(w)$ can be observed in figure 11.

Figure 11 shows the kernel density estimate of the non-migrant wage (log of the actual wage divided by the quarter average) distribution (solid line) already calculated in figure 3 together with the counterfactual density (dashed line) corresponding to the wage emigrants should be earning according to their observable characteristics. As a result, the difference between the two densities reflects the part of selection that is due only to observable characteristics of the migrants. The rest of the difference with the actual wage distribution of the emigrants can be considered as the effect of unobservables in selection. This can be

²⁵The logit regresses the migration dummy from the ENET on the same variables used in Chiquiar and Hanson (2005): schooling groups, age, age squared, marital status and interactions of these variables with the schooling groups. The results of this auxiliary regression are available from the author upon request.

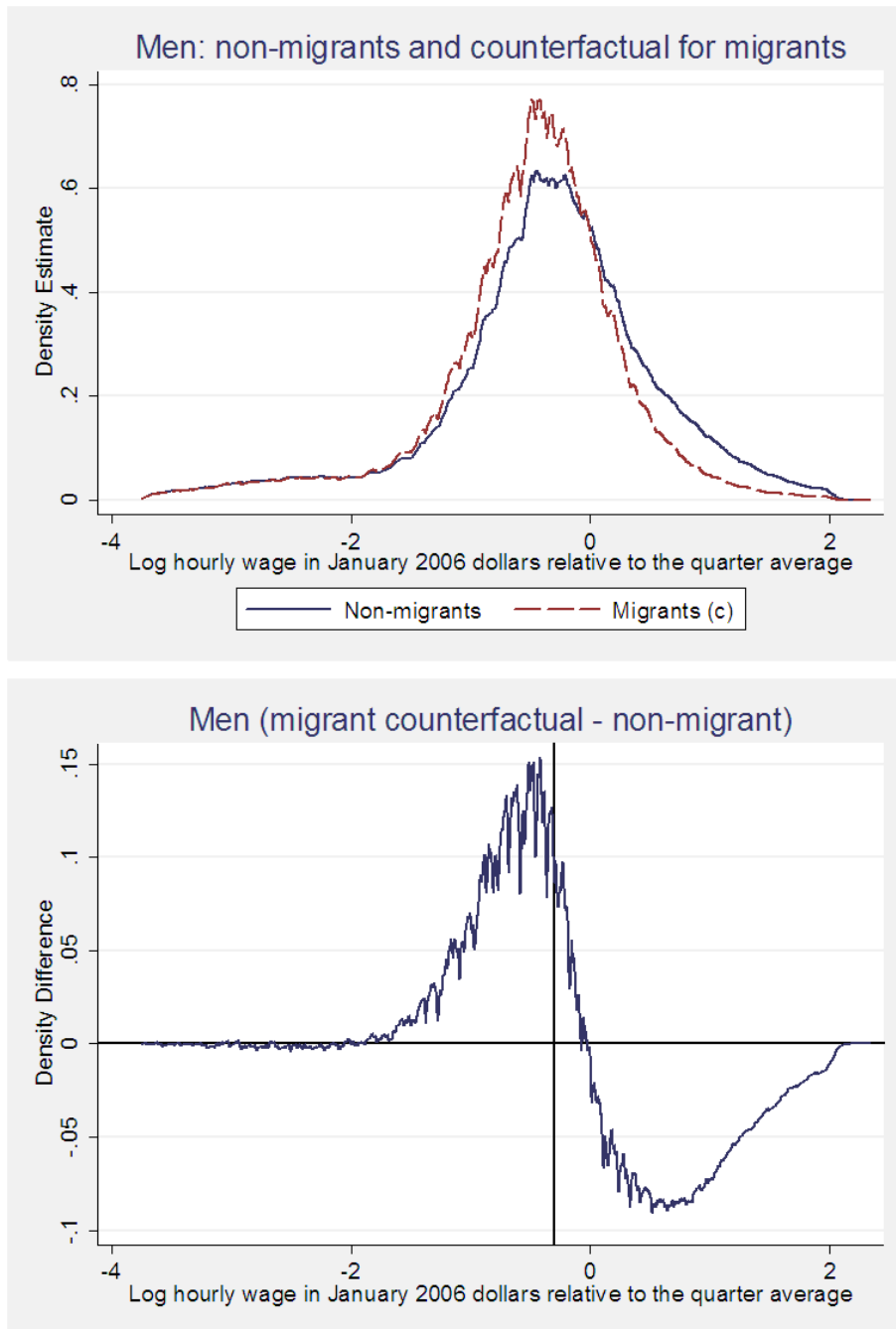


Figure 11: Wage distribution for non-migrants and counterfactual wage distribution for migrants

Source: ENET. The counterfactual is estimated following DiNardo, Fortin, and Lemieux (1996) and assuming that observations on wages for migrants are not available in the ENET.

quantified by computing the averages of the distributions represented in figures 3 and 11. The degree of selection (as defined in section 2) is -0.26 with the full ENET information (figure 3). The corresponding number for figure 11 is -0.16 so it can be concluded that 62 per cent of the negative selection result can be attributed to the effect of observables whereas the remaining 38 per cent corresponds to the effect of unobservables in selection. It must be observed that both types of characteristics (observable and unobservable) go in the same direction: negative selection.

The existence of negative selection in unobservables could be attributed to a variety of reasons. It could be the case that those who decide to migrate are those who receive negative wage shocks right before the migration move. For one fifth of the sample, it is possible to explore the evolution of wages during the four quarters preceding the migration move. When this is done, there is no downward bias in the wage during the year previous to the migration decision. Of course, it could still be the case that the Ashenfelter dip happens at an earlier period but that cannot be explored with this dataset. Other explanations for the negative selection in unobservables could be low unobserved ability or worse access to local networks by those who decide to emigrate.

Coming back to the comparison with Chiquiar and Hanson (2005) result, figure 11 also determines that the methodology alone cannot fully explain why they obtain positive instead of negative selection. The next step is to study whether the difference in data sources can.

5.1.2 Using US Sources to Obtain Data on Immigrants

A second hypothesis about why the results in this paper differ from Chiquiar and Hanson (2005) has to do with the quality of the data or, more exactly, with the ability of US sources to capture a fair representation of the migration flow from Mexico. Since the last US Census corresponds to 2000, the exercise will use instead the American Community Survey (Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2004)) from 2000 to 2004.²⁶ Table 2 shows that the relevant observable characteristics (age and schooling) for recent (arrived the previous year) Mexican immigrants in the United States do not differ much in the US Census and the ACS for 2000. They are also comparable to the general statistics for

²⁶The ACS has more observations on immigrants than the Current Population Survey. Disregarding the difference in time periods and using the data on Mexican immigrants from the US Census instead of the ACS leads to analogous results.

Summary Statistics (US sources on recent Mexican immigrants)				
Individuals aged 16 to 65		US Census 2000	ACS 2000	ACS 2000-2004
<i>Percent Male</i>		62%	66%	64%
		(0.00)	(0.03)	(0.01)
<u><i>Men</i></u>				
<i>Age</i>				
	Average	26.6	27.2	28.0
		(0.09)	(0.89)	(0.30)
	Median	24	24	25
<i>Schooling years</i>				
	Average	8.6	8.9	8.9
		(0.04)	(0.33)	(0.12)
	Median	9	9	9
<u><i>Women</i></u>				
<i>Age</i>				
	Average	28.0	28.9	29.9
		(0.13)	(1.45)	(0.45)
	Median	25	25	27
<i>Schooling years</i>				
	Average	8.8	8.9	9.2
		(0.05)	(0.52)	(0.16)
	Median	9	9	9
<i>Observations</i>		21,930	244	2,658

Table 2: Summary Statistics (US sources on Mexican immigrants)

Source: Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2004). Standard errors in parentheses and in smaller font. The 2000 US Census observations correspond to Mexican born individuals from the 5% Census sample who arrived in the United States in 1999. The American Community Survey (ACS) observations correspond to Mexican born individuals who were still residing in Mexico a year before the survey.

the period 2000-2004, used in the analysis below.

The next step is to replicate Chiquiar and Hanson (2005) estimation but using data for non-migrants from the ENET and data from recent Mexican immigrants to the United States from the ACS. Formally, this amounts to re-estimating $\hat{g}_{M,0}(w)$ but substituting the ACS immigrants for the ENET emigrants. The result is shown in figure 12.²⁷

Figure 12 represents the known distribution of wages of non-migrants (figures 3 and 11) but this time together with the counterfactual wage distribution representing how much Mexicans who have been in the United States for a year would earn, were they to come back to

²⁷The auxiliary logit regression is calculated as in footnote 25. Results available from the author upon request.

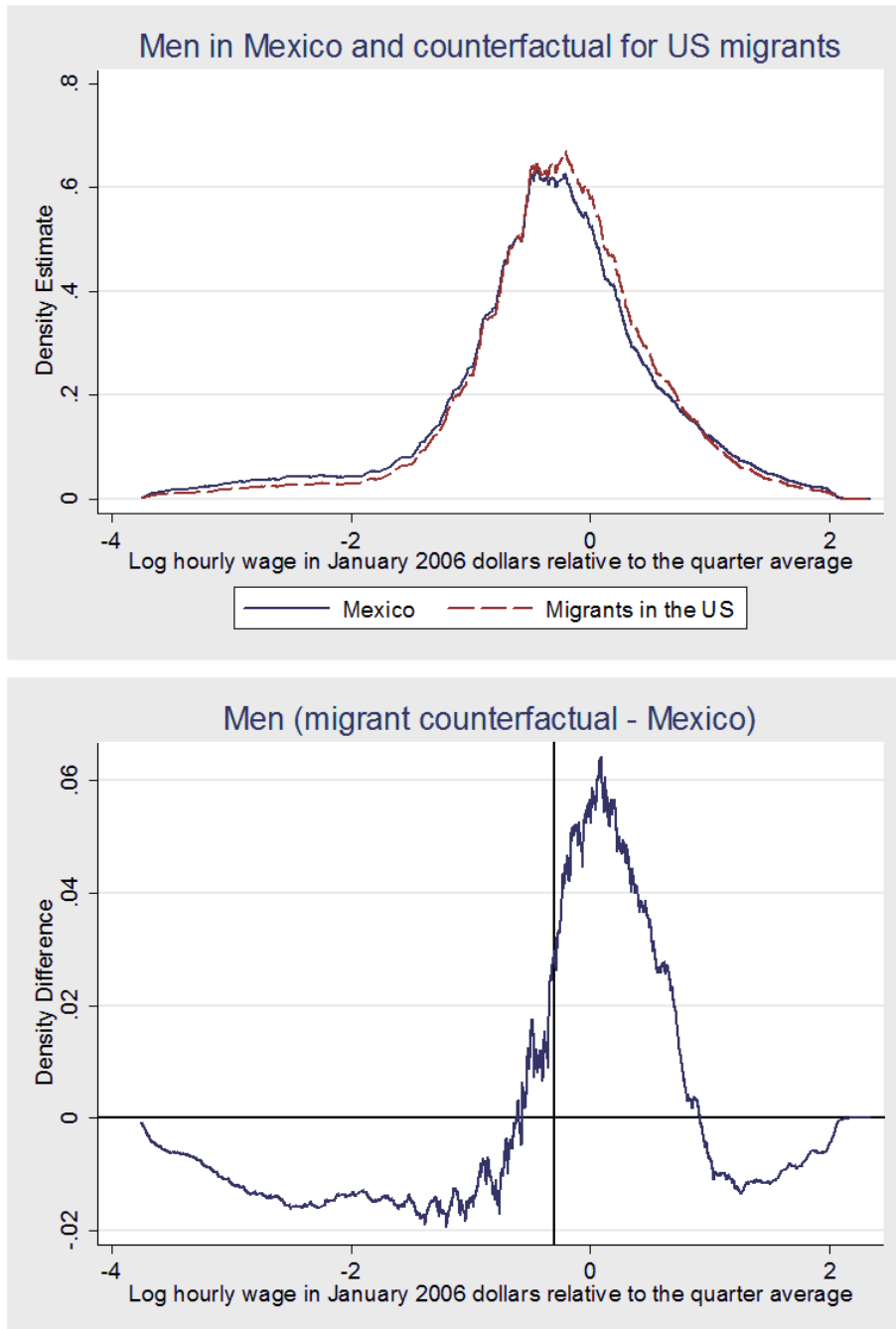


Figure 12: Wage distribution for non-migrants and counterfactual wage distribution for Mexican immigrants in the US

Source: ENET and ACS. The counterfactual is estimated following DiNardo, Fortin, and Lemieux (1996) and it can be interpreted as the wages Mexican immigrants in the US (arrived the last year) would earn were they to return to Mexico.

Mexico, according to their observable characteristics. This counterfactual wage distribution lies slightly to the right of the actual wage distribution for non-migrants, which would be evidence of positive selection. In other words, we can replicate Chiquiar and Hanson (2005) result for the migrant stock with the closest concept of flows available in US sources. As discussed above, this is a valid concept of flows as long as temporary migrants, who go from Mexico to the United States and return within a year, are not significantly different from those who stay for more than a year and are thus supposed to be enumerated by US sources. If this was the case, the reason for the difference between my result and that of Chiquiar and Hanson (2005) would not lie in the different time period or the methodology but only on the quality and characteristics of the data. Again, this statement can be quantified. The degree of positive selection implicit in figure 12 is calculated to be 0.07, contrasting with -0.26 in the first panel of figure 3. Taking the number for figure 11 (-0.16), this implies that the difference of results between Chiquiar and Hanson (2005) replication and this paper can be attributed 30 per cent exclusively to the omission of unobservables and 70 per cent to data differences.

In this respect, Chiquiar and Hanson (2005) were aware that the under-count of immigrants in the US Census was a potential problem for their result and this is the reason why they produced robustness checks in which they imputed data from the Mexican Migration Project to the estimated amount of under-counted. The reasoning was that, given that the MMP migrants come from rural communities in Central Mexico and many of them are reported to be undocumented, they could be a good representation of the immigrants not captured by the US Census. The next subsection explores how selection in the MMP compares to selection in the ENET.

5.2 Comparison with the MMP

The use of the MMP data was insufficient to overturn Chiquiar and Hanson (2005) positive selection result. In fact, the MMP dataset itself presents evidence of positive selection, as shown by McKenzie and Rapoport (2007a) and Orrenius and Zavodny (2005). Both papers estimate the probability of emigration on several controls and find that it is increasing and then decreasing in the schooling level of the individuals in the sample, which is evidence of intermediate to positive selection in schooling.

It turns out that even this result is consistent with the ENET dataset. The MMP surveys

individuals in rural communities of traditional emigrant sending Mexican states (mostly in Western and Central Mexico). Since the ENET is nationally representative, special regions or subsets of the Mexican population can be selected to observe selection inside these groups. Dividing the ENET sample by regions or states does not change the negative selection result. In fact, selection is negative in all regions and states in Mexico during the period except for the East of the country, where the selection result is ambiguous.²⁸

The special dimension of the MMP that leads to the positive selection result is its rural nature, defining as rural the population living in localities with less than 2,500 inhabitants. Figure 13 presents the kernel density estimate of the wage distributions for migrants and non-migrants within the Mexican rural population.

This representation offers now a picture that is different from what was seen in figure 3. In rural Mexico, which, as it can be seen in table 1, accounts for little more than a fifth of the Mexican population but more than two fifths of its emigrants, these are positively selected, with those earning higher wages and thus more productive emigrating more than individuals with a low wage.

To sum up, the fact that Mexican emigrants are negatively selected in terms of their wage stemming from the analysis of the ENET is perfectly consistent with the seemingly contradictory findings of previous literature. Studies based on US sources lack fundamental data on undocumented migrants whereas studies based on the MMP risk extrapolating to urban Mexico a behavior that turns out to be idiosyncratic to rural Mexico.

In terms of the best theory to study selection, the negative selection result was consistent with the simplest version of Borjas (1987) model. However, the result in rural Mexico shows that a model with heterogeneous migration costs decreasing in productive characteristics can be more appropriate to study the migration decision. One particular theory of heterogeneous migration costs decreasing in productive characteristics that can be tested with the ENET data is the existence of credit constraints in the migration decision. A hypothesis to test in future work is whether the difference in selection patterns between rural and urban Mexico can be explained by the existence of credit constraints in rural but not in urban areas of Mexico.

²⁸Results available from the author upon request.

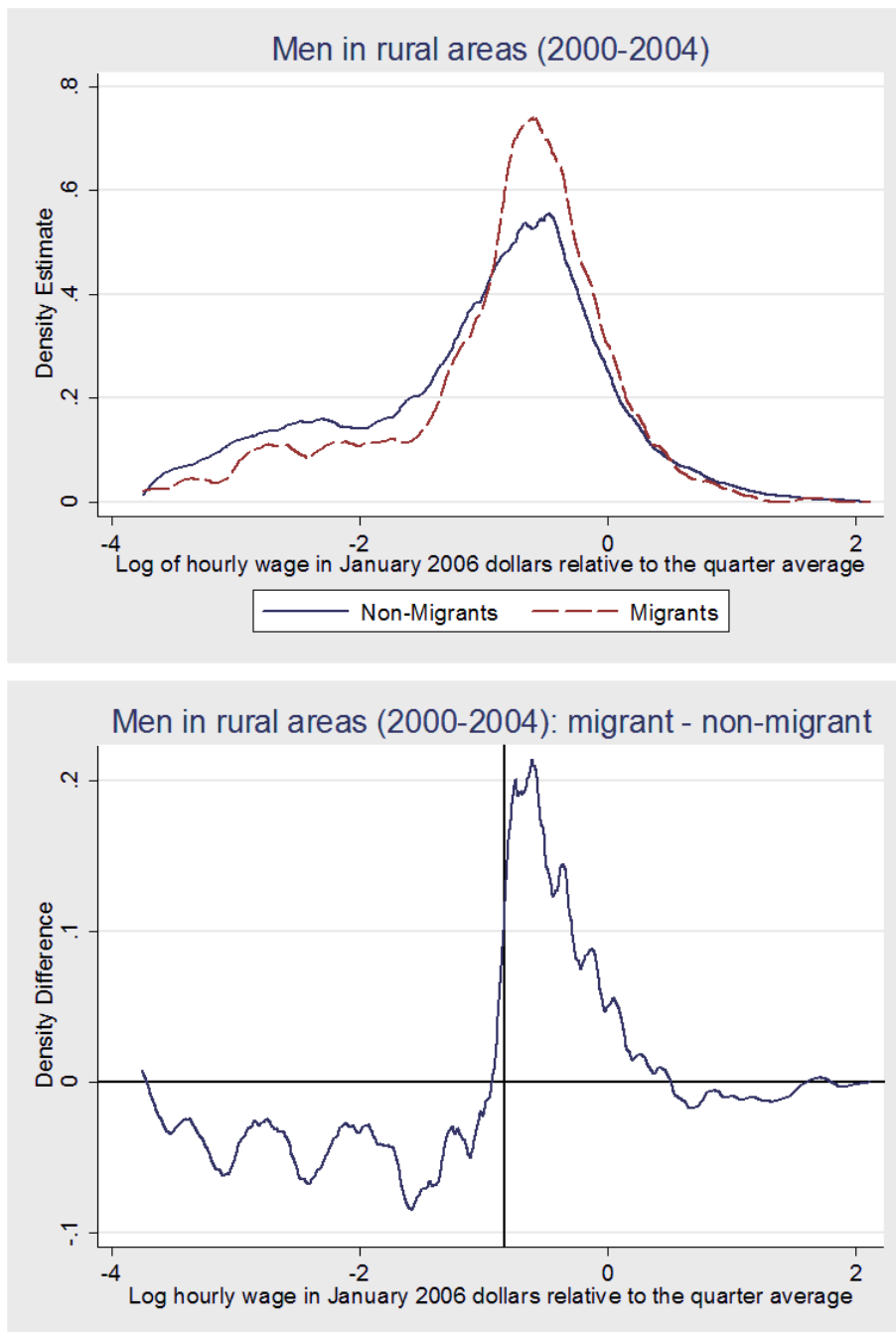


Figure 13: Wage distribution for migrants and non-migrants in rural areas

Source: ENET. See figure 3 for notes on the kernel density estimation. Rural population defined as those living in localities with less than 2,500 inhabitants.

6 Conclusion

This paper studies the productive characteristics of people who choose to emigrate internationally. The residents of an emigrant sending country are interested in knowing whether their more productive individuals are those who are leaving whereas the residents of an immigrant receiving country want to know how those who enter their country will affect them. Emigrant selection is said to be negative if the productive characteristics of the people who leave are on average lower than the productive characteristics of the people who remain at home.

Selection theory shows that both positive and negative selection are possible outcomes, depending on how the migration decision is modeled. As a result, it is an empirical question to determine which of these assumptions are more reasonable and whether there exists positive or negative selection in the emigration flow between a host and sending country. In this paper, the emigration between Mexico and the United States between 2000 and 2004 is the particular case studied thanks to the use of the ENET dataset, which offers the possibility of comparing directly non-migrants and emigrants right before they take the decision to leave their country. However, the methodology used here can surely be applied to other countries. Non-parametric techniques are extensively used in this paper in order to observe the distribution of characteristics of migrants and non-migrants.

The empirical results show clearly the existence of negative selection in Mexico in the period 2000-2004. Mexican emigrants to the United States are on average less productive in terms of their wage levels than non-migrant Mexicans. They are also younger and less educated. This is described in section 4.

The reasons why this main result differs from the existing literature are developed in section 5. The structure of the ENET dataset allows one to replicate Chiquiar and Hanson (2005) estimation techniques. A fundamental difference between their paper and this paper is the fact that they measure selection in the stock of Mexican immigrants in the United States whereas the ENET measures selection in the flow of emigrants as they leave the country. In order to address this issue, I use data on recent Mexican immigrants (arrived a year earlier) in US sources to replicate Chiquiar and Hanson (2005) positive selection result. It could still be the case that very short term migrant flows would entirely drive the negative selection outcome in this paper but the absence of seasonality in the degree of selection seems to indicate that the most important source of discrepancy is the fact that US sources

do not adequately enumerate Mexican migrants. In addition, it is possible to show that the methodology they used, based on DiNardo, Fortin, and Lemieux (1996), is also responsible for part of the discrepancy. The methodological problem is not severe because the effect of unobservables on selection goes in the same direction as the effect of observables. Even when the effect of observables (seniority, education, etc.) is discounted, there is still negative selection of emigrants. This negative selection in terms of unobservables could be attributed to a variety of factors: bad luck of future migrants, low unobserved ability, less access to networks in the original country, etc.

A different subset of previous literature which had also obtained positive selection results was based on the study of the Mexican Migration Project (MMP). The MMP studies rural communities in traditional emigrant-sending Mexican states. The ENET provides evidence that is consistent with this subset of the literature. Selection is indeed positive in rural Mexico (localities with less than 2,500 inhabitants), also according to the ENET. The conclusion is that the negative selection result found for Mexico as a whole is driven by urban Mexico. The reasons why rural and urban Mexico show different selection patterns are the subject of future research although one hypothesis that can be tested with the ENET data is whether credit constraints might affect the migration decision in rural Mexico more than in urban Mexico.

In general, it is relevant to understand who the emigrants are and also who they will be in order to assess the impact they can have on their origin and destination countries. In this sense, this type of dataset would also be interesting to study emigration flows among other countries.

Bibliography

- ANGELUCCI, M. (2005): "U.S. Border Enforcement and the Net Flow of Mexican Illegal Migration," *IZA Discussion Paper*, 1642.
- BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77(4), 531–553.
- (1990): "Self-Selection and the Earnings of Immigrants: Reply," *American Economic Review*, 80(1), 305–308.

- (1999): “The Economic Analysis of Immigration,” *In Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, 3A.
- (2003): “The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration in the Labor Market,” *Quarterly Journal of Economics*, 118(4), 1335–1374.
- BORJAS, G. J., AND B. BRATSBERG (1996): “Who leaves? The Outmigration of the Foreign Born,” *Review of Economics and Statistics*, 78(1), 165–176.
- BORJAS, G. J., AND L. F. KATZ (2007): “The Evolution of the Mexican-Born Workforce in the United States,” *In Mexican Immigration to the United States*, edited by George J. Borjas.
- CAPONI, V. (2006): “Heterogeneous Human Capital and Migration: Who Migrates from Mexico to the United States?,” *IZA Discussion Paper Series*, 2446.
- 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. R. (1978): “The Effect of Americanization on the Earnings of Foreign-Born Men,” *Journal of Political Economy*, 86, 897–921.
- (1999): “Immigration Policy and Immigrant Quality. Are Immigrants Favorably Self-Selected?,” *American Economic Review*, 89(2), 181–185.
- (2007): “Are Immigrants Favorably Self-Selected? An Economic Analysis,” *In Migration Theory: Talking Across the Disciplines*, edited by Caroline D. Brettell and James F. Hollifield, Second Edition.
- CONAPO (2006): “Aspectos Metodológicos,” Consejo Nacional de la Población. Secretaría de Gobernación de México.
- CUECUECHA, A. (2005): “The Characteristics of the Immigrants from Mexico to the US,” *Working Paper, ITAM-CIE*.
- DHS (2004): *2004 Yearbook of Immigration Statistics*. Department of Homeland Security, <http://uscis.gov/graphics/shared/statistics/yearbook/index.htm>. Web accessed on 1-26-2006.

- DiNARDO, J., N. M. FORTIN, AND T. LEMIEUX (1996): “Labor Market Institutions and the Distribution of Wages. 1973-1992: A Semiparametric Approach,” *Econometrica*, 64, 1001–1044.
- GATHMANN, C. (2004): “Effects of Enforcement on Illegal Markets: Evidence of Migrant Smuggling at the Southwestern Border,” *IZA Discussion Paper*, 1004.
- HANSON, G. H. (2006): “Illegal Migration from Mexico to the United States,” *Journal of Economic Literature*, 44(4), 869–924.
- HANSON, G. H., AND A. SPILIMBERGO (1999): “Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the U.S.-Mexico Border,” *American Economic Review*, 89(5), 1337–1357.
- IBARRARÁN, P., AND D. LUBOTSKY (2007): “Immigration and Self-Selection: New Evidence from the 2000 Mexican Census,” *In Mexican Immigration to the United States*, edited by George J. Borjas.
- INEGI (2004): *Módulo sobre Migración 2002. Encuesta Nacional de Empleo*. Instituto Nacional de Estadística, Geografía e Informática. México. Web accessed on 2-20-2006: www.inegi.gob.mx.
- (2005): *Encuesta Nacional de Empleo Trimestral (2000-2004)*. Instituto Nacional de Estadística, Geografía e Informática. Aguascalientes, México.
- LACUESTA, A. (2006): “Emigration and Human Capital: Who Leaves, Who Comes Back and What Difference does it Make?,” *Documento de trabajo Banco de España*, 0620.
- LEIBBRANDT, M., J. LEVINSOHN, AND J. MCCRARY (2005): “Incomes in South Africa since the fall of Apartheid,” *NBER Working Paper Series*, 11384.
- MARCELLI, E. A., AND W. A. CORNELIUS (2001): “The changing profile of Mexican migrants to the United States,” *Latin America Research Review*, 36(3), 105–131.
- MCKENZIE, D. J., AND H. RAPOPORT (2007a): “Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico,” *Journal of Development Economics*, 84(1), 1–24.

- (2007b): “Self-selection patterns in Mexico-U.S. migration: The role of migration networks,” *CREAM Discussion Paper*, 01/07.
- MISHRA, P. (2007): “Emigration and wages in source countries: Evidence from Mexico,” *Journal of Development Economics*, 82(1), 180–199.
- MUNSHI, K. (2003): “Networks in the Modern Economy: Mexican Migrants in the US Labor Market,” *Quarterly Journal of Economics*, pp. 549–599.
- ORRENIUS, P. M., AND M. ZAVODNY (2005): “Self-selection among undocumented immigrants from Mexico,” *Journal of Development Economics*, 78, 215–240.
- OTTAVIANO, G. I., AND G. PERI (2006): “Rethinking the Effects of Immigration on Wages,” *NBER Working Paper Series*, 12497.
- PASSEL, J. S., AND R. SURO (2005): *Rise, Peak, and Decline: Trends in U.S. Immigration 1992-2004*. Pew Hispanic Center. Washington D.C.
- REYES, B. I. (1997): *Dynamics of Immigration: Return Migration to Western Mexico*. Public Policy Institute of California. San Francisco, CA.
- REYES, B. I., H. P. JOHNSON, AND R. VAN SWEARINGEN (2002): *Holding the Line? The Effect of the Recent Border Build-up on Unauthorized Immigration*. Public Policy Institute of California. San Francisco, CA.
- ROBERTSON, R. (2000): “Wage Shocks and North American Labor Market Integration,” *American Economic Review*, 90(4), 742–764.
- ROY, A. D. (1951): “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, 3, 135–146.
- RUGGLES, S., M. SOBEK, T. ALEXANDER, C. A. FITCH, R. GOEKEN, P. K. HALL, M. KING, AND C. RONNANDER (2004): *Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor]. Web accessed on 7-15-2006: <http://www.ipums.org>.
- SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

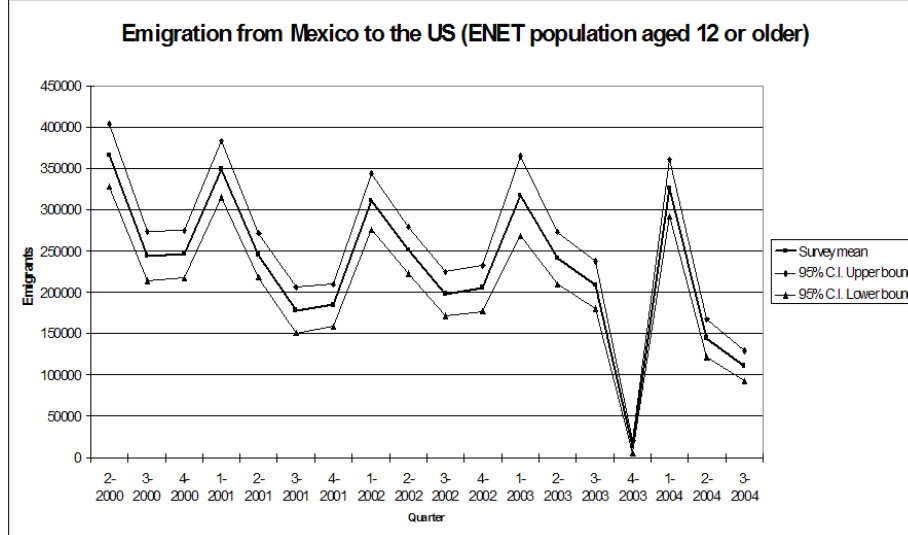


Figure 14: Emigration from Mexico to the US in the ENET (by quarter)

Source: ENET.

SJAASTAD, L. A. (1962): “The Costs and Returns of Human Migration,” *Journal of Political Economy*, 70(1), 80–93.

WORLD BANK (2006): *Global Economic Prospects. Economic Implications of Remittances and Migration*. Washington, D.C.

A The ENET Data and its Comparison with Other Sources

This appendix describes more thoroughly the ENET dataset on which this paper is based. For a first look at the data, Figure 14 shows the estimated number of migrants captured by the ENET between the second quarter of 2000 and the third quarter of 2004 for the population aged 12 or older.²⁹

Since this source is not designed to study migration but unemployment³⁰, it is relevant to have a look at the standard errors imposed by the survey design. This can be observed

²⁹The estimated migration rates for the population of less than 12 years old imply an average emigration of 51,015 children per year.

³⁰Unemployment in Mexico is 1.8 per cent in 2000-2004 (see table 3), which is about four times larger

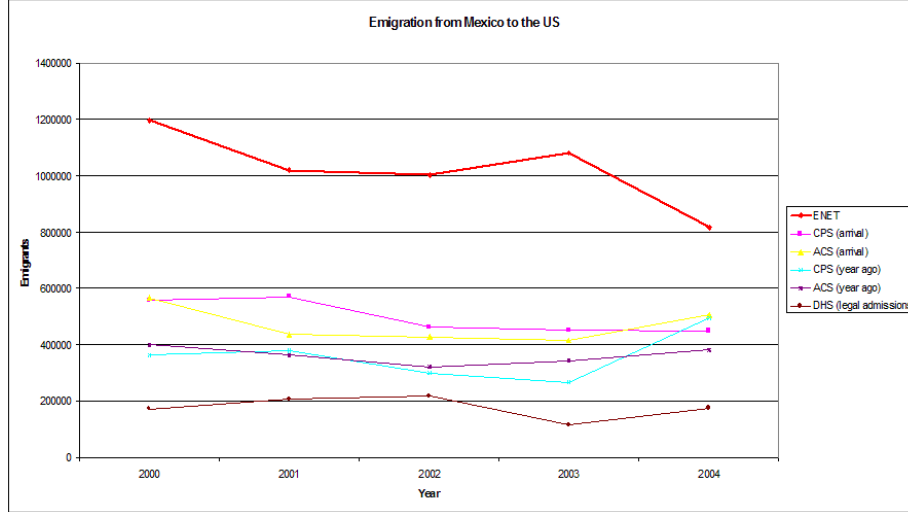


Figure 15: Emigration from Mexico to the US (different sources)

Source: Passel and Suro (2005) for the CPS and ACS series and DHS (2004) for legal admissions numbers. The ENET series accumulates quarterly migration numbers for 2001 and 2002. The 2000, 2003 and 2004 number is obtained by averaging through the available quarters and adding the average to the remaining quarter, which leads to a likely underestimation of flows for 2000 and overestimation for 2003 and 2004.

in the 95 per cent confidence intervals for the number of emigrants. There are ups and downs in the series that seem to correspond to seasonal variation in the number of emigrants (Hanson and Spilimbergo (1999)), clearly higher in the first quarter of the year (between 300,000 and 350,000 emigrants) and lower in the third and fourth quarter (around 200,000), with the second quarter somewhere in between. The anomaly in the fourth quarter of 2003 corresponds to a dataset problem (see footnote 20). Only the last two quarters of the series (second and third quarter of 2004) present a substantial reduction in emigration flows.

The yearly migration figures suggested by the ENET greatly diverge from those found in American sources, such as the ones documented in Passel and Suro (2005) and reproduced for comparison purposes in figure 15.

The reason for this great divergence may come from the different concepts these sources are measuring. American sources capture a subset of Mexican migrants who are in the United States at the moment of the survey and who declare when they entered the country. The than the average migration rate. Since the survey is designed to provide reasonable standard errors for the unemployment rate calculation, the corresponding standard errors for the migration rate could be too big to draw meaningful conclusions.

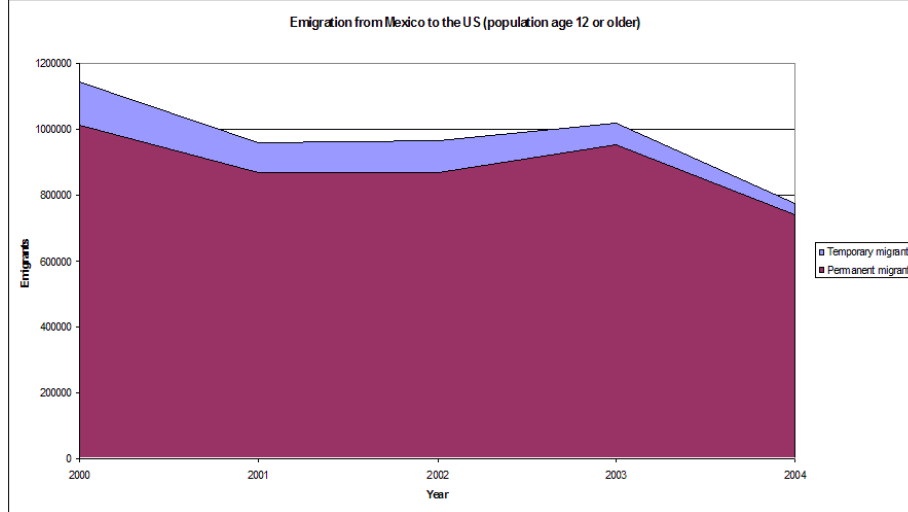


Figure 16: Emigration from Mexico to the US (temporary and permanent migrants)

Source: ENET. Definitions of temporary and permanent migrants taken from the questionnaire and described in the text.

ENET reflects Mexican emigrants who are not in Mexico at the time of the survey but who might come back later in the year or even make several trips to the United States. The ENET questionnaire allows one to break the migrant series between temporary and permanent migrants to check how important this discrepancy might be. The resulting distribution can be observed in figure 16.

The distinction between temporary and permanent migrants in the ENET depends on the amount of time an individual has been absent from the household and whether it plans to return in a period of less than three months. Many temporary migrants end up becoming permanent migrants whereas individuals initially defined as permanent migrants may decide to come back. The two groups are considered jointly for all the analysis in the paper (breaking the data between temporary and permanent migrants does not affect the results though). Those classified as temporary migrants represent, on average, 9 per cent of total migration to the United States per year during the period. With these caveats in mind, the point estimate from the ENET for permanent migration to the US in the year 2000 is one million Mexicans, which is still off the numbers in most of the sources cited by Passel and Suro (2005) and reported in figure 15. However, this figure is not that far from Passel and Suro (2005) estimate from the 2000 US Census of 750,000 Mexicans entering the US in 2000. More precisely, it is as far as the other estimates Passel and Suro (2005) report from the ACS and

CPS, only in the sense of over-estimating rather than underestimating. Unfortunately, there is no Census information to compare the data for the following years but this suggests that the Mexican source can be as valuable a measure of the size of Mexican emigration as the traditionally used American sources, which may suffer from under-counting recently arrived undocumented immigrants (Hanson (2006)).

Next, a comparison is made with traditional sources for the study of migration, such as the MMP (see footnote 5 in the text), the EMIF (see footnote 17 in the text) or the American Community Survey (ACS; Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2004)). This is done in terms of the characteristics of Mexican migrants in table 3 for the year 2000, following Hanson (2006), who compares the MMP with the Mexican and US Census for 1990.

It is apparent that the ACS under-samples males relative to Mexican sources in the same way that the US census does relative to the Mexican census (Ibarrarán and Lubotsky (2007)). Males represent between 80 and 90 per cent of immigrants in Mexican sources whereas they only account for 67 per cent in the ACS. The usual reason given in the literature is the fact that US sources under-sample undocumented aliens, who are supposed to be mostly men. Neither the ENET nor the ACS asks for the legal status of migrants but both the MMP and the EMIF do. The MMP estimates that only 22 per cent of Mexican males who migrated in the period 2000-2004 had proper documents whereas the estimate for females is 42 per cent. Although the sample size is small (878 men and 210 women), the hypothesis that men tend to travel undocumented more than women cannot be rejected at a five per cent significance level. However, the analysis in the EMIF yields different results. The point estimate is still lower for men since only 41 per cent of those who tried to cross the border by land from 2000 to the second quarter of 2003 claimed to carry legal documents to enter the US for 46 per cent of women but this time the null hypothesis that both estimates are equal cannot be rejected at a 95 per cent confidence level even though the sample size is bigger than the MMP's (9,471 men and 895 women).

To sum up, it is not clear that the under-sample of undocumented immigrants explains why relatively more women appear on the ACS than in Mexican sources. Another hypothesis is that Mexican sources gather information about both temporary and permanent migrants whereas American sources would only capture information from more permanent migrants. The ENET seems to support this hypothesis since there is a higher proportion of men in

Migrant Mexican Nationals, Males

		ENET 2000*	ACS 2000	MMP 2000	EMIF 2000*	
					Attempted	Captured
Percent Male		82%	67%	86%	90%	83%
Age	18 to 27	48%	63%	44%	26%	52%
	28 to 37	27%	19%	32%	38%	33%
	38 to 47	15%	12%	19%	24%	13%
	48 to 57	8%	2%	5%	8%	2%
	58 to 67	3%	3%	1%	4%	0%
Males aged 18 to 47						
Years of Schooling	None	3%	2%	2%	4%	4%
	1 to 4	16%	8%	13%	15%	11%
	5 to 8	38%	32%	44%	44%	49%
	9 to 11	31%	16%	26%	32%	27%
	12 to 15	9%	38%	12%	4%	8%
	16 plus	3%	3%	3%	1%	0%
Live in urban area		58%			49%	52%
In Labor Force		91%	86%	100%	79%	81%
Work in agriculture		37%	10%	15%	27%	22%
Sample size, 18-67		2,277	147	309	1,329	3,953
Sample size, 18-47		2,055	137	292	1,244	3,863

Migrant Mexican Nationals, Females

		ENET 2000*	ACS 2000	MMP 2000	EMIF 2000*	
					Attempted	Captured
Percent Female		18%	33%	14%	10%	17%
Age	18 to 27	58%	51%	58%	29%	68%
	28 to 37	21%	23%	28%	38%	23%
	38 to 47	10%	12%	12%	11%	7%
	48 to 57	6%	6%	1%	16%	2%
	58 to 67	4%	6%	0%	5%	0%
Females aged 18 to 47						
Years of Schooling	None	2%	3%	0%	3%	3%
	1 to 4	8%	9%	2%	17%	9%
	5 to 8	29%	25%	43%	42%	42%
	9 to 11	30%	15%	32%	16%	34%
	12 to 15	23%	37%	21%	10%	11%
	16 plus	8%	10%	3%	11%	1%
Live in urban area		68%			60%	51%
In Labor Force		37%	41%	50%	52%	38%
Work in agriculture		2%	13%	0%	2%	2%
Sample size, 18-67		578	83	73	121	934
Sample size, 18-47		504	73	71	96	911

* Last three quarters

Table 3: Characteristics of Mexican migrants in different sources

Source: ENET, ACS, MMP and EMIF. The EMIF Attempted category refers to individuals who declare they will try to cross to the United States without a return date, going to work or look for work or who plan to stay for more than a year. The EMIF captured category refers to individuals seized by the United States Border Patrol.

the subgroup of temporary migrants than in the one of permanent migrants for all of the studied years although it must not be forgotten that the group of temporary migrants only amounts to 9 per cent of total emigration.

Coming back to table 3, the age distribution for male migrants appears similar in all sources, with the age group from 18 to 27 years old concentrating approximately half of the population of immigrants, which is consistent with the view of migration as a long term investment whose reward is higher for individuals of young age. Only the EMIF attempted migrants have a lower concentration in this segment, probably due to the stratification by age of its questionnaire.

Females follow the same pattern in their age distribution with, if anything, an even higher concentration of young migrants (over 50 per cent in most sources).

Table 3 also shows the schooling distribution for individuals aged 18 to 47 years old. The highest concentration of male migrants occurs in the category from 5 to 8 years of education according to Mexican sources (from 38 per cent in the ENET to 49 per cent in the captured EMIF), that is, primary school completed or almost completed plus several years of high school (primary school is finished after six years of schooling). This range is not far from the 32 per cent recorded in the ACS for this group. In fact, the ACS schooling distribution for migrants is remarkably similar to the one found in Mexican sources except for the fact that the mass of individuals between 9 and 15 years of schooling is split differently. For example, the ENET gathers 40 per cent whereas the ACS sums up to 54 per cent. The biggest difference is how this numbers split between the two categories: 9 to 11 years old and 12 to 15 years old. The first one has 31 per cent in the ENET for 16 per cent in the ACS whereas the second counts 9 per cent in the ENET and 38 per cent in the ACS. This is the reason why the average schooling years for Mexican males in the ENET is 7.3 whereas the ACS gets a result of 9 years. Abstracting from the differences in coverage of different surveys signaled above, Ibararán and Lubotsky (2007) hypothesize that Mexicans in the US might tend to over-report their true education level. The Mexican high school system has two levels, the first one of which ends after three years (Secundaria) with nine years of schooling. It could be the case that people finishing the “Secundaria” claim to have finished high school in US sources.

There is even more disagreement among sources for the female schooling distribution although migrant women appear to be more educated on average than migrant men. The

top category of schooling is still 5 to 8 years in the MMP and EMIF surveys but goes up to the 9 to 11 years group in the ENET.

Another relevant statistic in table 3 is labor force participation. It is higher for males in the ENET and MMP (91 and 100 per cent) than in the EMIF (79 per cent for attempted and 81 per cent for captured migrants) due to the fact that EMIF migrants are surveyed after leaving home whereas ENET migrants are surveyed before leaving and MMP respondents have to remember what they were doing when they left. Female labor force participation is notably lower in all sources, from 37 per cent in the ENET to 52 per cent in the EMIF attempted migrants category. ACS numbers are not comparable since they refer to labor force participation once the immigrants are already in the United States. However, it is interesting to notice that the labor force participation behavior of Mexican immigrants is similar in the United States to the one they have at home: 86 per cent for males and 41 per cent for females.

Finally, Hanson (2006) also includes the percentage of Mexicans who work in agriculture and related industries in his comparison of different migration sources. The reason is that Mexican immigration into the United States was traditionally linked to the agricultural sector, especially during the Bracero Program (1942-1964) and it is interesting to check whether this remains true. According to the ACS, only 10 per cent of male Mexican migrants aged 18 to 47 years old entering the United States in 2000 worked in the agricultural sector (13 per cent of women did). According to the ENET, 37 per cent of male Mexican emigrants used to work in the agricultural sector before leaving. This number is higher than but comparable to both EMIF surveys. However, it is surprisingly higher than the MMP number of 10 per cent, especially taking into account that the MMP is over-representing rural communities in Central and Western Mexico. The explanation could be in the small sample size for the MMP since the result is based on just 29 observations, contrasting with 2,055 observations for the ENET.