HAL
open science

New insights into the selection process of Mexican migrants.What can we learn from discrepancies between intentions to migrate and actual moves to the U.S.?

Isabelle Chort

To cite this version:

Isabelle Chort. New insights into the selection process of Mexican migrants.What can we learn from discrepancies between intentions to migrate and actual moves to the U.S.?. 2012. halshs-00689467

HAL Id: halshs-00689467

https://shs.hal.science/halshs-00689467v1

Preprint submitted on 19 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2012 – 18

New insights into the selection process of Mexican migrants.What can we learn from discrepancies between intentions to migrate and actual moves to the U.S.?

Isabelle Chort

JEL Codes: F22, J61, O15

Keywords: Migration ; Roy model ; Migrant Self-Selection ; Subjective Data ; Shocks ; Mexico

PARIS-JOURDAN SCIENCES ECONOMIQUES
48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ECOLE NORMALE SUPÉRIEURE – INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

New insights into the selection process of Mexican migrants. What can we learn from discrepancies between intentions to migrate and actual moves to the U.S.?*

Isabelle Chort†

February 9, 2012

Abstract

Comparing intentions to migrate and actual migration of Mexicans, I intend to assess the impact of unexpected shocks and misevaluated costs on the materialization of migration plans. I show that both sets of reasons may explain discrepancies between intentions and subsequent actions without denying the rationality of intentions by resorting to the theoretical framework of the Roy model. I use intention and migration data from the Mexican Family Life Survey, together with precipitations monthly series, hurricane and crime data to represent different sets of shocks. Correlations between intentions and migration on the one hand, and between intentions and individual labor market characteristics show that intentions are not devoid of informational content. Then, modelling intentions and migration with a bivariate probit, I find that shocks, and in particular rainfall and hurricanes, affect the probability to migrate conditional on initial intentions. The key finding is nonetheless the much lower propensity for women to migrate abroad conditional on intentions, which suggests that women incur specific costs or constraints misestimated at the intention stage. Alternative explanations, such as gendered preferences are discussed, but convergent empirical evidence suggest that women are more constrained than men on the international migration market. Moreover the data suggest that migrant are positively selected with respect to their unobserved characteristics whereas those with intention to migrate abroad are negatively selected. The shift in selection between the two stages of the migration process may be due to the cost reducing effect of individuals' unobserved characteristics that explain their higher local wages.

Keywords: Migration, Roy model, migrant self-selection, subjective data, shocks, Mexico

JEL codes: F22, J61, O15

*I am grateful to Sylvie Lambert, Philippe De Vreyer, Alice Mesnard, François Bourguignon, Elisabeth Sadoulet, and participants in various seminars and conferences for helpful comments.

†Paris School of Economics. Address: 48, Boulevard Jourdan - 75014 - Paris (France), Phone: +33 1 43136364, E-mail: chort@pse.ens.fr

1 Introduction

The selection of migrants is a core issue of the analysis of migration and its impacts. Indeed, migrants differ from non migrants in both observable and non-observable ways. This paper aims at providing new insights into the selection mechanism of emigrants by exploring together intentions to migrate and migration decisions of Mexican adults.

The analysis of Mexican international migration is interesting on two accounts. On the one hand, though emigration from Mexico to the United States has been extensively documented and explored, since the founder works of Borjas (1987), no consensus has emerged concerning the nature of the selection of Mexican migrants. On the other hand, the specific design of the Mexican Family Life Survey (MxFLS) can be exploited in an original way in order to shed a new light on the selection of Mexican migrants. Panel data from the first two waves (2002 and 2005-2006) of the MxFLS project are currently available. Surveyed individuals being asked in the first wave of the survey whether they have the intention to migrate, their answers can be compared with their subsequent behavior. I focus in this article on potential inconsistencies between stated intentions and observed actions.

They could first be due to the nature of intention data: since they do not imply any commitment one could consider them as mere noise. The degree to which subjective data can be trusted is under debate in the literature. This interpretation is nonetheless ruled out in the empirical study, by the positive and significant correlation between intention and actual migration. The emigration rate is for example more than twice higher among individuals having the intention to move. This point is further debated in the following sections.

If one accepts to consider the value of intention data, numerous questions arise: What explains that among those who wish to migrate only a small fraction has actually moved three years later ? Are actual migrants different from those who did not carry out their migration plans within this three year span? If so, how to theoretically account for a two-stage selection process ? And how are individuals selected at each step of the migration decision ?

The originality of this article is to consistently study in a single theoretical framework both subjective – intentions – and objective – migration behaviors – data, even though they happen to be conflicting at first glance. Section 3 adapts to this aim the self-selection model developed by Roy (1951), which has been first applied to migration by Borjas (1987) and has now become

a reference framework for the analysis of migrant selection (Borjas and Bronars (1991), Chiquiar and Hanson (2005), Fernandez-Huertas Moraga (2008)). Indeed, economic agents are considered in the following theoretical developments to first choose to participate in the local (Mexican) labor market, or to migrate to the United States. At this initial stage, they state an intention to move or stay. But whereas intentions do not necessarily encompass all present and future constraints to the achievement of individual migration plans, actual migrations have to occur in the real world. The absence of materialization of early plans could be explained first by the fact that individuals may be hit by unexpected shocks between their intention statement and their actual migration, that would lead them to reconsider their decision given their new environment. Second, although rational, individuals may not be able to accurately estimate all costs linked to migration at the intention stage, because of the complex nature of these costs and available information. In that respect, it is shown here that the size and even the sign of the selection of Mexican migrants, depending on both observable and unobservable characteristics can differ whether costs are accurately estimated or not, and depending on the specification of the cost function. Finally, candidates to migration could be forced to postpone their plans due to liquidity constraints, as assumed by (Borger, 2011). We take this hypothesis into account by specifying a cost function general enough to allow for restricted credit access. According to this interpretation, shocks may have an additional effect on the materialization of migration intentions by affecting individuals' capacity to fund their travel, either positively (unexpected gain, inheritance) or negatively (for example adverse weather shocks consuming savings intended for migration).

Then, the empirical part of this study first aims at providing elements supporting a rational interpretation of intentions. Second, I estimate jointly two equations for the probability to have the intention to migrate and actually migrate with a bivariate probit, allowing for a correlation of unobserved individual terms in both equations. Such an empirical model seems indeed best appropriate to make allowances for unexpected shocks, miscalculated costs or liquidity constraints to explain migration conditional on intention to move, which are the main explanations suggested by the theoretical discussion, while bringing to light the nature of selection on observable individual and household characteristics at both stage of the migration decision.

Different sets of shocks variables are constructed, based in particular on global gridded

datasets gathering long term series of rainfall data. Shocks are found to affect migration, consistently with recent findings on Mexico-U.S. migration measured on the U.S. side by Pugatch and Yang (2011), but they appear to be only part of the story: the largest shift in selection between intentions and migration is related to gender. Conditional on initial intentions, women are found to be much less likely to migrate than men. As suggested in the theoretical model, the constraints linked to gender might have been underestimated by female would-be emigrants. Alternative interpretations, such as gendered preferences are discussed in section 4.

Finally, I provide an estimation of the selection of individuals on the basis of their unobserved characteristics at both stages of the migration process. I find suggestive evidence of a slight shift in selection on unobservables, from negative at the intention stage to positive for actual migrants, consistently with Borger (2011). However, such a shift is insignificant relative to results on gender.

The paper proceeds as follows: In section 2, I first briefly present some basic facts on Mexican migration and review the literature on the selection of Mexican migrants. I then discuss the relevance of an economic analysis of intentions, and more specifically intentions to migrate. Section 3 recalls the main intuitions of the model described by Roy (1951), adapting it to the analysis of the migration decision represented as a two-stage process. Section 4 first provides an overview of the data; Then, bivariate probit regressions are used to analyse the impact of shocks and other individual and household characteristics on the materialization of early migration plans, with a particular focus on gender issues. In addition, I estimate the sign of self-selection relative to unobservables at each migration stage. Section 5 concludes.

2 A Review of Literature

2.1 What does migrant selection mean?

The word selection is sometimes ambiguous. First, migrant selection in a strictly econometric sense only means that migrants are not a random sample of the population of their home country. Attempts to estimate any effect of migration on a range of outcomes, either on migrants themselves or on their household of origin thus require technical procedures to deal with this selection bias. Except in the very rare situations where a randomized experiment can be exploited (McKenzie, Gibson, and Stillman (2009a); McKenzie, Gibson, and Stillman (2009b)),

correction for selection relies on statistical procedures (Wooldridge, 2002), preferably backed by the use of adequate instruments. But since migration is at the heart of the economy of the household, appropriate instruments that could correct for migrants' selection are usually difficult to find (McKenzie and Sasin, 2007). Migration decisions, allocation of the labor force, educational choices, decisions of expenditures and investments are necessarily and definitely the most tightly entangled.

Nonetheless, a large stream of literature has been interested in exploring the nature of migrant selection as regards a number of more or less desired characteristics from a destination country's viewpoint since Borjas (1987). Consequently, selection is described as either positive or negative, depending on a comparison between migrants and non-migrants based on observable characteristics, among which is for example education.

Moreover, studies relying on host-country data result in blurring the distinction between selection due to external constraints and self-selection (Liebig and Sousa-Poza, 2004). If this distinction may seem somewhat artificial, it can be nonetheless argued that migrant samples studied thanks to host-country data are shaped at least as much by immigration policies as by migrant self-selection. Having access to intention data would thus provides a way to tackle the origin of the selection process, although individuals may of course take at least partially into account immigration policies of potential destination countries at the intention stage.

A brief summary of previous findings in the literature on the selection of Mexican migrants follows a presentation of some basic facts on Mexican migration.

2.2 Mexican migration and migrant selection

The massive scale migration from Mexico to the U.S. is now a hundred-year-old phenomenon. Though extensively studied and documented (see for example Durand, Massey, and Zenteno (2001) for a review of literature), Mexican migration flows are not easily quantified. Estimates of the size of the population of Mexican migrants in the U.S. were between 7.0 and 7.3 million in 1996, among which 2.3 to 2.4 were unauthorized (Binational Study on Migration, 1997). Ten years later, the same population is estimated to have reached 11.5 million (American Community Survey 2006 data, cited in Batalova (2008)).

Migration outflows from Mexico are tightly linked to the economic and historical develop-

ments of its northern neighbor. The origin of Mexican migrations to the U.S. dates back to the turn of the twentieth century and the low skilled labor intensive building of railroads connecting Mexico provinces to the U.S.. Recruitment of labor in the nearest densely populated Western states of central Mexico initiated a process that was to affect deeply both economies. After a break caused by the recession of the 1930s, migration flourished during the Second World War, within or outside the legal framework of the Bracero program aimed at supplying the U.S. with agricultural workers, and during the boom period of the sixties and seventies (Freeman and Bean, 1997). The flows of Mexican migrants to the U.S. continuously increased from the sixties to the end of the nineties, even though the number of illegal migrants may be underestimated: The number of Mexicans entering the U.S is estimated to 315,000 per year during the period 1990-1996, whereas it hardly reached 290,000 during the whole period 1960-1970 (Binational Study on Migration, 1997).

In the recent period, restrictive immigration policies were implemented, symbolized by the building of a wall on the frontier, while at the same time the sustained demand for flexible unqualified labor in U.S. farms led authorities to close their eyes on the presence of undocumented workers (Cornelius (2001), Hanson (2006)). Although it is still too soon to assess the effects of recent legal restrictions to immigration in the U.S. and the consequences of the 2008 still ongoing crisis, both have no doubt resulted in increased migration costs, at least for illegal migrants[1].

Summarizing existing data from different sources[2], Durand, Massey, and Zenteno (2001) find much continuity over time in migrants' geographical origin and individual characteristics: Mexican migrants are males who migrated relatively young and predominantly originate from the western states of Mexico with a strong migratory tradition, though evidence of new origins for Mexican migrants arise at the end of the nineties. They find no tremendous change in

[1]As regards our period of interest (2002-2005), costs may have also been increasing: in a recent lecture at the Norface Migration Conference (London, 6-9 april, 2011), Douglas Massey presented estimates of smuggler's fees which suggested at least a 30% increase between 2002 and 2005. This point will be further discussed when interpreting the results in section 5.

[2]The main surveys of the past three decades are the Mexican Migration Project, the ENADID and EMIF surveys. The Mexican Migration Project initiated in 1982 has been surveying each year a few communities and the U.S. destination areas of the migrants originating from these selected communities. The ENADID survey (Encuesta Nacional de la Dinámica Demográfica) (counting three waves in 1992, 1997 and 2006) is a large scale household survey (nearly 58,000 surveyed household in the first wave) that gives a representative picture of migrants to the U.S. along with members of their household of origin left in Mexico, but excludes settled migrants from its scope. Adopting an approach centered on flows, the EMIF survey (Encuesta sobre Migracion a la Frontera Norte) selects migrants' samples at a few points of the border. Thirteen waves were conducted at regular intervals since 1994.

migration patterns but a decrease over time in the relative educational skills of migrants, after controlling for cohorts effects.

On this very issue of selection of Mexican migrants as regards observable (and in particular education) and unobservable skills, no consensus has emerged. Following Borjas (1987), the study of migrant selection on their unobservable characteristics relies on the assumption that wage reflects productivity. The share of the wage which remains unexplained after controlling for classical determinants (in particular education and experience) is interpreted as the return on migrants unobservable skills.

Among recent papers, and with different data sources Ibarraran and Lubotsky (2005) and Fernandez-Huertas Moraga (2008) find evidence of negative selection of Mexican migrants, which is consistent with Borjas' hypothesis. Note that the former focuses on education, whereas the latter considers both observable and unobservable skills. Nevertheless, Ibarraran and Lubotsky (2005) admit that illegal migrants may be undercounted in the census data they exploit. And in fact, using survey data from the INEGI, Fernandez-Huertas Moraga (2008), finds that selection is positive for migrants from rural areas. On the contrary, Orrenius and Zavodny (2005) in line with Chiquiar and Hanson (2005), and resorting to data from the Mexican Migration Project find that Mexican undocumented migrants are selected in the middle of the income distribution. Caponi (2006), on the other hand, proposes a U-shape selection, with Mexicans with the highest and lowest levels of education being more likely to migrate. McKenzie and Rapoport (2007) reconcile some of these conflicting results by linking the nature of selection to the size of migration networks, implying that migrant selection varies over time in communities involved in migration. They show that in Mexico the probability to migrate increases with education in communities with few or no networks and decreases with education in communities with large networks. In the same line, the recent paper by Borger (2011) formalizes the cost-decreasing impact of networks, and focuses on liquidity constraints resulting in intermediate selection, under the assumption that returns to skills of Mexicans in the U.S. are decreasing. However, his empirical results of intermediate selection at the beginning of the 1990s are based on a durable consumption index, and do not refer to education nor migrants' unobservable skills. As regards unobservables, his results suggest a positive selection of migrants during most of the period he studies (2000-2009), although wage residuals are found to be much volatile.

Concerning selection relative to gender, one of the first attempts to specifically address the issue of Mexican women's migration is due to Donato (1993). Basing on household models for migration she argues that "in Mexico, households adopt a strategy of sending at least one member, usually the male household head or son, to work in the United States", because of the "cultural beliefs and traditional values about the roles of women and men in families". She indeed finds evidence of family migration, women's migration being conditioned by the prior successful migration of a husband or father.

Quoting a more recent study by Kanaiaupuni (2000) who uses data collected by the Mexican Migration Project (1999), women's lower probability to migrate may be due to traditional roles devoted to women in the family organization:

> First, norms associated with the role of women with children limit the social acceptability of migration among mothers. Second, greater demands imposed by children impede the geographical and job mobility often required of migrants. Third, the lower costs of raising a family in Mexico than in the U.S. tend to encourage split household migration strategies. Hence, married women with children are likely to remain in the sending communities while male family members migrate (Kanaiaupuni 1998)"

Note however that she finds no evidence of a negative correlation between having children and the probability to migrate, contrary to her hypotheses.

2.3 Intentions to migrate

Intentions are part of a larger set of subjective data which are often mistrusted by economists (Bertrand and Mullainathan, 2001). Indeed in a strict revealed preferences approach, intentions are irrelevant and only actions matter. On the other hand, recent developments in psychology have brought subjectivity, and thus intentions at the forefront of investigations. The notion of intention is indeed subjective, but inextricably linked to action. In order to understand why people do not act as they said they would, the theories of reasoned action (TRA) (Ajzen and Fishbein, 1980) and planned behavior (TPB) (Ajzen, 1991) elucidate conditions for intentions to be predictive of behaviors (Sutton, 1998). These theories suggest that intentions are unconstrained, whereas constraints appear in the transition to the domain of action.

In a closely related field, geographers much investigated migration and residential mobility (Cadwallader, 1992), but kept their attention focused on internal migrations in developped countries and, though they present both objective and cognitive determinants of migrations, the gap between them is not filled. At a crossroads of psychology and geography, Lu (2005) studies inconsistencies in behaviors considering residential mobility, referring to TRA and TPB as well as to the concepts of "blocking of mobility" and "unexpected moves" (Moore, 1986). Using more than 40,000 observations from the American Housing Survey Data in the 1985-89 period, he stresses the role of the demographic composition of the household and past moving experiences among the determinants of residential mobility behaviors.

In a recent paper, Van Dalen and Henkens (2008) focus on discrepancies between intentions to migrate and international migration in the Netherlands. They specifically question the quality of intentions as predictors of actions regarding migration. Categorizing individuals according to their intentions and actions, they study two consistent types (stayers and movers) and an inconsistent one, which they name dreamers, made of individuals having the intention to migrate but who did not move [3]. They show that dreamers do not substantially differ from movers and conclude to the satisfying quality of intentions as predictors of actions, in view of an emigration rate of 24% in a two-years time among individuals with migration plans. Nonetheless, financial and legal constraints are no doubt much higher for would-be emigrants originating from developing countries.

Data on intentions to migrate are even more valuable when applied to selection issues. Indeed, as emphasized by Liebig and Sousa-Poza (2004), following Chiquiar and Hanson (2005) most studies on immigrant self-selection are based on destination country data. As a consequence, self-selection and selection due to immigration policies of the host country are inextricably mixed. Liebig and Sousa-Poza (2004) insist on the usefulness of data on individual intentions since they would capture migrant self-selection at the source, free from host country specific influence. This assertion may be discussed, as noted above, since intentions may incorporate individuals' understanding of immigration policies in host countries. However, the relevance of intentions data is undeniable if they can be compared to actual migration data. The choice made by Liebig and Sousa-Poza (2004), citing Burda et al. (1998), to assume that intentions

[3] They do not consider the symmetrical inconsistent case (intention to stay and migrate) in their paper due to the small number of observations in their overall sample (300 individuals).

are "a monotonic function of the underlying driving variables which motivate migration" is thus debatable. Being here able to compare intentions to subsequent actions I claim that intentions are reliable, informative and rational, but they cannot be considered suitable proxies for actual migration. Indeed, in the particular case of Mexican migration, using intention to proxy for actual migration would lead to a large overestimation of women's migration.

At this point, it is necessary to introduce a theoretical structure in order to provide a better understanding of the selection of Mexicans into migration, represented as a two-stage process.

The following section is based on the model of self-selection presented by (Roy, 1951). The original contribution of this article on a theoretical ground lies in the emphasis put on the discussion of the specification of costs and its impact on migrant selection.

3 Modelling intention to migrate and migration: the Roy model reinterpreted

In this section, I adapt the classical Roy model to the analysis of intentions to migrate. The key underlying assumption is that intentions are indeed rational. This point could be debated but empirical evidence shown in the next section do not challenge the rationality assumption. Then, two sets of explanations are proposed in order to account for discrepancies between intentions and actions, still in a rational setting.

3.1 Intentions

The Roy's model of selection (Roy, 1951; Heckman, 1974 ; Heckman and Sedlacek, 1985) has become the classical theoretical framework to analyse the migration decision since it has been adapted by Borjas (1987; Borjas and Bronars, 1991).

This study proposes a reinterpretation of the Roy model in order to explain individuals' intentions to migrate. I thus draw upon the above cited authors, as well as on Magnac (1991) whose application of the Roy model is more specifically related to labor market issues, to recall the main assumptions and results of the Roy model, adapted to migration intentions.

Consistently with neoclassical migration theories (Todaro, 1969; Sjaastad 1962) intentions to migrate are considered to be stated on the basis of a comparison of potential earnings in the origin and destination countries. Though initial mobility choices, in a model derived from

Roy's, are assumed to be based on a comparison of wages, the model can be easily extended to take into account unemployment (see Appendix).

Leaving aside unemployment issues, in a simplified framework, individuals are here assumed to choose between two options: migrate to the U.S. or stay in Mexico[4]. It comes down to considering that individuals self-select either in the local (Mexican) labor market or in the foreign one (U.S.). They are assumed to make a choice on the basis of a comparison of their utilities in each case, and utility is further assumed to depend on wage only, and not on non-monetary characteristics of jobs on each labor market.

Following Magnac (1991), wages on each market can be decomposed as follows:

$$\begin{cases} \text{U.S.:} & \ln(w_{\text{U.S.}}) = X\beta_{\text{U.S.}} + t_{\text{U.S.}} \\ \text{Mexico:} & \ln(w_{\text{Mex}}) = X\beta_{\text{Mex}} + t_{\text{Mex}} \end{cases} \tag{3.1}$$

Where X is a vector of variables representing the part of the wage determined by observable characteristics of individuals (such as education, experience) as in a Mincer-type wage equation. $\beta_{\text{U.S.}}$ and β_{Mex} represent returns to observable individual skills on each market. $t_{\text{U.S.}}$ and t_{Mex} are the components of wages explained by unobservable individual skills, though they are known with certainty by each individual (Magnac, 1991).

Since it is much more costly to migrate than to participate in the local Mexican labor market, migration costs are included in the model. Note that an equivalent way of regarding the cost differential associated with the participation on each market, is to consider that Mexicans are rationed on the U.S. labor market: the adequate model is thus a segmented (Magnac, 1991) or generalized Roy model.

The founder articles of Borjas (1987) and Borjas and Bronars (1991)) considered constant time-equivalent migration cost, for more tractability. This approach was criticized by Chiquiar and Hanson (2005) who assume that migration costs decrease with schooling. In a labor market perspective, Magnac (1991) allows costs to vary with both observable (education, experience) and unobservable (ability) individual characteristics. But since he considers that costs are

[4]Internal migrations are not taken into account here. Indeed, the costs to internal migration are likely to be much lower than the costs to migrate out of Mexico, and the changes in migrant selection due to the introduction of costs in the model is precisely one of the points of this section. Moreover the additional simplification that consists in limiting the set of destinations to the United States is not that much unrealistic given empirical evidence

determined by the supply-side, he does not allow them to depend on individual non market characteristics.

In the particular context of migration, Sjaastad (1962) separates money costs, that he defines as "expenses incurred by migrants in the course of moving" and non-money costs among which are opportunity costs and what he names psychic costs[5]. In addition to Magnac (1991)'s cost function, and consistently with the migration literature, costs are here allowed to also depend on a number of variables that do not enter the wage equations such as proxies for the opportunity costs of migration (having a secure job in the formal sector for example), but also proxies for psychic costs. The latter include demographic characteristics of the household (being married or having young children), but also previous individual moves, since in the definition proposed by Sjaastad (1962), psychic costs are supposedly higher for individuals with no prior moving experience[6]. Variables controlling for household wealth and assets should also be considered in the cost function since economic characteristics of the household affect migrants' ability to take care of some of these costs. Finally, as is emphasized by Carrington, Detragiache, and Vishwanath (1996), networks are expected to affect migration costs. Consequently, I define the following cost function:

$$C_i = c + X_i\gamma_0 + Z_i\gamma_1 + \eta_i \tag{3.2}$$

where c is a basic fixed cost, X is a vector of human capital variables that determine wages in equations (3.1), Z is a vector of non-market variables that do not enter wage equations[7]. Borrowing Magnac's image, η_i accounts for individual ability to position oneself in the queue formed by rationed individuals trying to get to the foreign labor market, in the same way as the $t_{\text{U.S., Mex}}$ in the wage equations, which means that though unobservable this term is known with certainty by each individual.

Note that the formulation of the cost function chosen here is general enough to include liquidity constraints. Indeed, as emphasized by Borger (2011), liquidity constraints may well constitute the principal obstacle for candidates to emigration to carry out their plans, if they

[5] "Since people are often genuinely reluctant to leave familiar surroundings, family, and friends, migration involves a 'psychic' cost"(Sjaastad, 1962).

[6] This latter variable was not included in all empirical specifications since it is highly endogenous

[7] See the empirical section for a list and definition of all variables included in Z.

cannot have access to credit. The author argues that migrant networks could substitute for deficient credit institutions by lending money to help migrants to pay travel costs (including smugglers' fees). This assumption is taken seriously and examined in the empirical section by controlling for family and community network variables. Moreover all regressions include measures of household wealth and location dummies, likely to proxy for individuals' access to credit and banking facilities, in order to assess the potential role of liquidity constraints on the materialization of migration plans.

The relevant wage to be taken into account is thus the discounted-by-cost log wage $\ln(\tilde{w}_{\text{U.S.}}) = \ln(w_{\text{U.S.},i}) - C_i$. Agents are then expected to compare $\ln(\tilde{w}_{\text{U.S.}})$ and $\ln(w_{\text{Mex}})$. Under all previous assumptions, the individual choice rule writes:

$$\begin{cases} \text{Intention to migrate:} & \text{if} \quad \ln(\tilde{w}_{\text{U.S.}}) > \ln(w_{\text{Mex}}) \\ \text{Intention to stay:} & \text{if} \quad \ln(w_{\text{Mex}}) > \ln(\tilde{w}_{\text{U.S.}}) \end{cases} \tag{3.3}$$

In that framework, once intentions are assumed to be rationally stated on the basis on a simple maximization, two major sets of reasons can account for individuals' deviation from their own program without rejecting the rationality assumption: shocks or omitted and/or miscalculated costs due to lack of information.

3.2 Actual migration

First, individual plans may be affected by unexpected and unpredictable shocks before they could be materialized. These shocks may modify either individual characteristics X (health, ability) or the economic environment, and thus affect expected earnings on each market through $\beta_{\text{U.S.}}$ and β_{Mex}. All else equal, a positive shock increasing the expected wage on the local market w_{Mex} is likely to cause emigration candidates to postpone or cancel their plans. Shocks may also affect migration costs (such as changes in immigration policies in the U.S.), or the ability to bear these costs. Among the latter category, one can think of the negative impact on savings of adverse weather shocks or natural disasters in the absence of credit and insurance markets.

Second, migration costs could have been miscalculated or omitted at the intention stage. If individuals are rational there is a priori no reason why they would omit costs at the intention stage, assuming that they have all relevant information. However, this latter assumption can

be partly relaxed for two main reasons. First, it can be argued that whereas some costs are indeed undoubtedly known with certainty ahead of time (this is the case in particular for the constant part of the cost function, and in general for all direct costs including monetary costs associated with travel, visa or smugglers' fees for example, even though they may be subject to variation before migration plans are carried out), on the contrary, other costs such as those called "psychic" by Sjaastad (1962) may be more difficult to estimate. In particular, under the realistic assumption that the cost function includes an individual unobservable component, η_i, it may be as difficult for rational agents as it is for the researcher to ex-ante identify the specific impact of individual characteristics such as being a woman or having dependent children on the cost of migration. It is very likely that basing on observation, individuals are able to build a rough estimate of migration costs with a relative imprecision when making allowances to costs affected by observed and known variables (the $X_i\gamma_0 + Z_i\gamma_1$ part of the cost function presented in equation 3.2) and costs due to unobservable individual characteristics such as motivation, noted η_i. In addition, it could be maintained that even though information is fully available, at an early stage of their migration plans, individuals would either be still in the process of collecting information, or unwilling to incur yet all information costs and would thus make their intention statement without necessarily resorting to all available information[8]. A slightly different interpretation is favoured by Borger (2011), who uses a different theoretical framework: if individuals are credit constrained, they may have stated intentions to move on the basis of a rational calculation, but not amassed yet enough money to pay for their travel when surveyed for the second time. As above mentioned, this hypothesis is tested in particular by controlling for household income: if credit constraints is to explain the duration of the stay of emigration candidates in the home country between the moment they had the intention to migrate and their actual departure, then, we expect individuals from richer households to be more likely to have migrated between 2002 and 2005 conditional on intention to migrate.

[8]In that case, one can obviously consider that returns on both markets might be similarly affected by imprecise estimation. However, it seems natural to assume that information on wage differentials between home and foreign countries is at the origin of migration intentions, and individuals may be assumed to be willing to acquire it as soon as at the intention stage. For that reason, and for simplification purposes, in the remaining part of this article, only costs are assumed to be potentially estimated with imprecision

3.3 Implications on selection of imprecise cost estimation

The aim of this theoretical exercise is to show that even in a very simple model, miscalculation of the parameters of the cost function at the intention stage can affect the nature of the final selection of migrants relative to observable and unobservable characteristics. Note that for practical purposes, the functional form is assumed not to be modified. The parameters only are assumed to be subject to estimation biases, from the individual's viewpoint. This is a very restrictive assumption, whose justification is to show that even slight errors made by individuals in the estimation of migration costs when planning to migrate may cause selection to be modified between intention and migration stages. In keeping with the notations used in the model and the arguments presented above, any parameter of the cost function (equation 3.2) c, γ_0, γ_1 may be wrongly estimated at the intention stage. Subsection 3.3.2 investigates the ensuing impact on selection on observables, focusing on the example of education. In addition, selection on unobservables may also be modified, and a rather extreme example of error in estimating the cost function, where the unobservable component of the wage function η_i would not be taken into account at the intention stage, is documented in section 3.3.3

3.3.1 Selection on observables

Consider first migrants' selection on the basis of their observable characteristics: At the intention stage, as it has already been noted, selection on the basis of individual observable characteristics, and particularly education and experience variables, included in X, depends on the difference in returns on both markets, $\beta_{\text{U.S.}} - \beta_{\text{Mex}}$, minus the cost parameter relative to individuals' labor market characteristics γ_0. The nature of selection thus depends in particular on the degree of transferability of immigrants' education and experience (Chiswick and Miller, 2009), and on the relation between education and migration costs. A plausible assumption concerning the latter is that migration costs actually decrease with education (see for example McKenzie and Rapoport (2007)). Education may indeed facilitate access to information, legal immigration resources such as visas, and reduces adaptation costs to a new environment. If the parameter γ_0 is wrongly estimated at the intention stage, the selection on education is expected to vary between intention and actual migration. In the particular case of imprecise estimation causing an attenuation bias, we would observe a higher level of education in the population of migrants

than in the population of emigration candidates, but the reverse case is conceivable as well.

Then, the parameter γ_1 accounting for costs associated with non labor market characteristics is also likely to be affected by estimation biases modifying the selection relative to gender, household demographics, etc. The empirical section of the paper focuses on these specific issue.

3.3.2 Selection on unobservables

Now turn to the selection process relative to the country of origin [9] based on individual unobservable characteristics. A possibility to discuss the selection of Mexican emigrants, as Borjas (1987) does, is to compare the mean of earnings that migrants would receive in Mexico had they not migrated, to the mean of Mexican earnings over the whole sample.

Based on the model presented earlier, define I to be an index function: intention to migrate occurs when $I^* > 0$:

$$\begin{aligned} I^* &= \ln(\tilde{w}_{\text{U.S.}}) - \ln(w_{\text{Mex}}) \\ &= (X_i(\beta_{\text{U.S.}} - \beta_{\text{Mex}} - \gamma_0) - c - Z_i\gamma_1) + (t_{\text{U.S.}} - t_{\text{Mex}} - \eta_i) \end{aligned} \tag{3.4}$$

Define $\nu = t_{\text{U.S.}} - t_{\text{Mex}} - \eta_i$, $\nu \rightsquigarrow N(0, \sigma_\nu)$. Assume further that $\eta \rightsquigarrow N(0, \sigma_\eta)$. Thus, under the normality assumption ($t_{\text{U.S.}} \rightsquigarrow N(0, \sigma^2_{U.S.})$, $t_{\text{Mex}} \rightsquigarrow N(0, \sigma^2_{Mex})$) a given individual has the intention to migrate with the probability:

$$\begin{aligned} P &= Pr[I = 1] \\ P &= Pr[I^* > 0] \\ &= Pr[\nu > -X_i(\beta_{\text{U.S.}} - \beta_{\text{Mex}} - \gamma_0) + c + Z_i\gamma_1] \\ &= 1 - \Phi(z) \end{aligned} \tag{3.5}$$

with $z = \frac{-X_i(\beta_{\text{U.S.}} - \beta_{\text{Mex}} - \gamma_0) + c + Z_i\gamma_1}{\sigma_\nu}$ and Φ being the normal cumulative distribution function. Note also ρ the coefficient of correlation between $t_{\text{U.S.}}$ and t_{Mex}, and $\rho_0 = \text{corr}(t_{\text{Mex}}, \eta)$

The expected wage in Mexico of Mexican migrants had they not migrated, conditional on

[9] Selection relative to the country of destination – the question of immigrants' performance in the U.S. compared with that of U.S. natives – is as well discussed by Borjas (1987) but is out of the focus of this study

their observable characteristics, is:

$$E(\ln(w_{\text{Mex}})|I^* > 0, X_i) = X_i\beta_{Mex} + \frac{\sigma_{\text{Mex}}\sigma_{\text{U.S.}}}{\sigma_\nu}\left[\rho - \frac{\sigma_{\text{Mex}}}{\sigma_{\text{U.S.}}} - \rho_0\frac{\sigma_\eta}{\sigma_{\text{U.S.}}}\right]\lambda(z) \tag{3.6}$$

Because of the cost function chosen to best represent migration costs, the formula is more complex than in Borjas (1987) who assumed constant costs. The specification of migration costs indeed directly affects selection on unobservables. Consider as an illustration the case where individuals are mistaken when estimating a cost function and omit for example the unobservable term of the cost function above denoted η_i.

The latent variable, noted I'^* would in that case write:

$$\begin{aligned} I'^* &= \ln(\tilde{w}_{\text{U.S.}}) - \ln(w_{\text{Mex}}) \\ &= (X_i(\beta_{\text{U.S.}} - \beta_{\text{Mex}} - \gamma_0) - c - Z_i\gamma_1) + (t_{\text{U.S.}} - t_{\text{Mex}}) \end{aligned} \tag{3.7}$$

And equation 3.6, would then give:

$$E(\ln(w_{\text{Mex}})|I'^* > 0, X_i) = X_i\beta_{Mex} + \frac{\sigma_{\text{Mex}}\sigma_{\text{U.S.}}}{\sigma_\upsilon}\left[\rho - \frac{\sigma_{\text{Mex}}}{\sigma_{\text{U.S.}}}\right]\lambda(z') \tag{3.8}$$

With $\upsilon = t_{\text{U.S.}} - t_{\text{Mex}}$, $\upsilon \rightsquigarrow N(0, \sigma_\upsilon^2)$ and $z' = -X_i(\beta_{\text{U.S.}} - \beta_{\text{Mex}})/\sigma_\upsilon$

By omitting the individual unobservable component in the cost function, we uncover the equation discussed by Borjas (1987). In that case, the sign of the selection would depend on the sign of $\rho - \frac{\sigma_{\text{Mex}}}{\sigma_{\text{U.S.}}}$. If this term is positive then the mean of emigration candidates' earnings in Mexico is above mean Mexican earnings: individuals intending to migrate to the U.S. are selected in the upper part of the distribution of Mexican earnings, and one would talk about positive selection. On the contrary, if $\rho - \frac{\sigma_{\text{Mex}}}{\sigma_{\text{U.S.}}} < 0$, the selection is negative[10].

On the other hand, the nature of the selection when the unobservable component of the cost function is included depends on the sign of $\rho - \frac{\sigma_{\text{Mex}}}{\sigma_{\text{U.S.}}} - \rho_0\frac{\sigma_\eta}{\sigma_{\text{U.S.}}}$, as is shown in equation 3.6.

[10]Borjas (1987) makes the plausible assumption that the individual unobservable components of earnings in each country are highly correlated, that is ρ is assumed to be close enough to 1. Thus, the sign of the selection entirely depends on the relative unequality of the distributions of earnings in Mexico and in the U.S. Under the above assumption of a sufficiently high correlation between unobservable individual abilities on both markets, a greater variance in the distribution of earnings (conditional on observables X), in Mexico than in the U.S. ($\sigma_{Mex} > \sigma_{U.S.}$), implies that migration intentions to the U.S. will concern individuals in the lowest end of the distribution of earnings in the home country ($E(\ln(w_{\text{Mex}})|I^* > 0) < E(\ln(w_{\text{Mex}}))$. Then the initial self-selection would be expected to be negative: low skilled Mexicans have greater incentives to migrate. Conversely, if $\sigma_{U.S.} > \sigma_{Mex}$ higher skilled Mexicans have a comparative advantage on the U.S. rather than on the Mexican labor market and the population of Mexican emigration candidates is positively self-selected.

The implications of the omission of the unobservable component of migration costs as regards the sign of the selection of migrants can now be discussed by comparing equations 3.6 and 3.8. Depending on assumptions made on ρ_0[11], and on the relative size of the ratios $\sigma_{\text{Mex}}/\sigma$U.S. and σ_η/σU.S. and ρ[12], selection may be modified and its sign reversed, because of the misspecification of the cost function.

Taking up the conclusion of Fernandez-Huertas Moraga (2008),

> 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.

Note however that unlike Fernandez-Huertas Moraga (2008) whose objective is to assess the nature of Mexican migrants self-selection, the point of the theoretical discussion presented here was to show that even in a very simple theoretical framework, selection may be modified if costs are wrongly estimated at the intention stage.

3.4 Empirical strategy

The first point of the empirical section is to provide elements supporting a rational interpretation of intentions to migrate. Second, I intend to make allowances for shocks and constraints in interpreting the discrepancies between migration plans and actual behaviors. Third, I provide an estimate of the possible shift in individual self-selection between the intention and migration stages.

In order to translate the intuitions provided by the theoretical framework provided above to study selection, and explain migration conditional on early intentions, I estimate the following

[11] ρ_0 may be plausibly assumed to be negative and sufficiently close to one, which means that the component of earnings in Mexico explained by individual unobserved heterogeneity and individual ability relative to migration are highly correlated: the higher the unobserved ability or motivation (explaining higher wage in Mexico), the lower the migration costs incurred. Remember that the symmetric assumption is made concerning the correlation between individual unobservable components of earnings on both markets (correlation positive and close to one).

[12] Note that costs have an additional effect on selection through $\lambda(z)$. Since by definition $\lambda(z)$ increases with z, for example, an increase in costs, either related to market or non market variables, which leads to a decrease of z, causes all else equal an attenuation of the selection process whatever its sign.

model, made of two equations estimated simultaneously with a bivariate probit:

$$\begin{cases} \text{Intention to migrate} : y_{1,i} = a_1 + b_1 X_i + c_1 Z_i + u_{1,i} \\ \text{International migration} : y_{2,i} = a_2 + b_2 X_i + c_2 Z_i + d_2 S_{i,j,k} + u_{2,i} \end{cases} \tag{3.9}$$

The two dependent variables are the probability to have the intention to migrate, $y_{1,i}$, and the probability to actually migrate $y_{2,i}$. Since individual unobservables are likely to affect both intentions and actual migrations, the bivariate probit specification is appropriate since it allows the errors terms $u_{1,i}$ and $u_{2,i}$ to be correlated. The sets of independent variables, X and Z, enter both equations, whereas S is specific to the migration equation. Consistently with the rational modelization of intentions presented above, I expect intentions and migration to be driven by individual labor market characteristics X. The set of labor market characteristics used in the next section includes age and age squared, and education dummies. Consistently with the notations used in the theoretical model, and for lack of a direct measure of migration costs, the vector of variables Z referring to non market variables that is likely to affect migration costs along with the variables in X is included in both equations. All the variables used in the empirical section are listed and described in Appendix. Note already that in the empirical specification Z contains gender and demographic characteristics of the origin household such as the household size, the presence of children or elderlies in the household, plus networks variables, regional dummies and economic controls (expenditures proxying for household income). The vector S, added to the migration equation only, contains variables representing different shocks likely to have affected individuals between 2002 and 2005 (rainfall shocks, hurricanes at the state level, natural disasters at the community level, as well as shocks affecting the household itself, natural disaster, death or illness). Shocks are noted $S_{i,j,k}$ in the following equation, subscripts referring to the household, locality and state level. In addition, some specifications take potential non-linearities into account by allowing shocks to enter non additively the migration equation, and shocks variables are interacted in particular with gender and education.

The bivariate probit specification allows in particular to compute marginal effects of the independent variables entering the model on the probability to migrate conditional on intention $Prob(y_{2,i} = 1 | y_{1,i} = 1)$, while allowing for the correlation of residuals in both equations.

Second, the theoretical part of the paper focuses on international migration for both tractabil-

ity and simplification purposes. Nonetheless since internal migration data are available in MxFLS they provide an interesting point of comparison. I thus estimate a bivariate probit for intentions to migrate within Mexico and actual internal migration. Indeed, internal migration is much less costly than migration to the United States, both as regards direct costs of migration (including travel costs, documents or visas, smuggler fees) and numerous indirect costs (among which social and economic integration, language, risk). In the internal migration case, both shocks and costs are thus likely to have a smaller influence on the materialization of migration intentions.

Finally, although the empirical part of this paper focuses on selection relative to individual observable characteristics, I use wage data to investigate selection on unobservables. Following Borger (2011) I thus estimate wage equations of the following type:

$$\ln(w_i) = \alpha_0 + \alpha_1 Exp_i + \alpha_2 Exp_i^2 + \alpha_3 Schooling_i + \alpha_4 Gender_i + \alpha_5 Region + u_i \qquad (3.10)$$

Where Exp denotes individuals' experience[13], $Schooling$ is the number of years of schooling, and $Region$ stands for regional dummies. In addition, urbanization is controlled for by including dummies for different rural and urban strata. $\ln(w_i)$ is the log of the hourly wage measured in 2002 (first survey wave). The unobserved individual characteristics are defined as the residual wage u_i, after controlling for predictable factors. In order to assess the sign and possible change of selection relative to unobserved characteristics, I compare the mean and distribution of residual wage of individuals depending on their intention to move in 2002 and their actual migration status in 2005.

4 Empirical analysis

4.1 Data

The data used in this article come from different sources: The intention and migration data come from a panel formed of the two waves of the Mexican Family Life Survey (MxFLS)[14]. In order to represent exogenous unexpected and unpredictable shocks, I resort to precipitations

[13]For lack of specific data on individual experience, this variables is defined as the potential experience, derived from age and the number of years of schooling.

[14]http://www.ennvih-mxfls.org/

data, data on hurricanes and crime figures. All sources are briefly described below. Summary statistics are presented in Appendix.

MxFLS is a nationally representative household survey, with a longitudinal structure (documented in Rubalcava and Teruel (2006) and Rubalcava and Teruel (2008)). Two waves of data collection have been conducted up to now, in 2002 and 2005[15].

During the first wave, 8,440 households (19,177 individuals aged 18-64) were surveyed, in 150 communities and 16 states, representing all regions of Mexico. Since my objective is to compare intentions to subsequent decisions, I had to restrict my sample to the 15,917 individuals (83% of the initial sample) who were present in the household at the time of the survey so that subjective data could be collected. Those who were for example temporarily out when their household was surveyed could not be included in my sample for lack of intention data. Tracking of households and individuals led to a high re-contact rate (over 92%), but the trail of 1,237 adults has been lost. Most of the results shown in the followings are thus based on a subsample of around 13,000 adults (aged 18-64), present in the household in 2002, with non missing migration intention data, non deceased and tracked in 2005. Since attrition is mostly due to migration, either internal or international, it raises a very important issue as concerns the validity of the results, and is thus addressed in section 4.

Intentions to migrate are collected in the first wave. Note that individuals are first asked a very general question about their migration intentions. Then they are asked precisions regarding the destination they would choose. Wording is known to be of first importance as concerns subjective data (Bertrand and Mullainathan, 2001). For that reason, the very general way the question is formulated[16] is particularly appropriate since a more precise question could result in a greater propensity for interviewees to censor themselves. Moreover, such a formulation is more likely to fit the representation of unconstrained intentions developed in the theoretical part of this article. The question is particularly vague concerning any time-limit for the realization of migration plans, and in particular does not imply that they should be carried out within the next three years. Nonetheless this limitation does not challenge the relevance of these data to address the issue raised in this article, even though the interpretation might be slightly different:

[15]The second wave was actually completed in 2006.

[16]The exact question is: "Ha pensado usted en irse a vivir en un futuro, fuera de la localidad/colonia en la que vive actualmente ?" which could be translated as "Have you ever thought of moving one day out of the locality you are now living in ?"

among those who had the intention to migrate, who are those who were able to (or may be forced to ?) carry out their plans within a relatively short span of time, when others could or would not ?

As concerns the definition of migrants, two categories are considered: international migrants, first, are individuals who either migrated abroad and returned between 2002 and 2005, or were currently abroad when their household was reinterviewed. Similarly internal migrants are individuals who moved to another Mexican locality between 2002 and 2005, and either stayed in their new place of residence or returned[17].

Data on weather shocks (precipitations) come from global gridded datasets produced by the University of Delaware's Center for Climatic Research[18]. Using monthly series available from 1949, I applied exactly the same strategy as in Pugatch and Yang (2011) and created state-level[19] yearly normalized rainfall variables (rainfall z-scores)[20].

Data on hurricanes are collected from the Coastal Services Center database of the National Oceanic and Atmospheric Administration[21] and crime data are taken from the "Justice in Mexico"[22]and "Seguridad Pública en México"[23] projects.

All variables used in the subsequent regressions are listed and described in Appendix. Basic summary statistics are also provided in appendix (table 14). The regional categorization of Mexican states is taken from Durand, Massey, and Zenteno (2001), and the label "historic region" refers to the Western states of Central Mexico which first experienced massive outmigration flows to the United States at the end of the nineteenth century and have developed a long history of migration since then. In all regressions observations are weighted using survey weights and standard errors are clustered by household, except in table 9 (locality clusters).

I mainly focus in the remainder on individuals with intention to move in 2002, and the

[17]I chose not to call internal migrants those individuals who moved whithin the same locality or within the most urbanized strata of the capital city of Mexico

[18]Full documentation is available at:
http://climate.geog.udel.edu/c̃limate/html_pages/Global2_Ts_2009/README.global_p_ts_2009.html

[19]For reasons of confidentiality, data permitting the geographical identification of localities in MxFLS (latitude, longitude, codes that could be compared to INEGI codes) are not made public. This is why I use state-level rainfall data, even though global gridded datasets offer much more precision.

[20]I first assigned grid points to states based on latitude and longitude coordinates, then summed up monthly data to obtain yearly rainfall variables and computed state-level averages for each year, state-level long term averages (1949-2005) and state-level standard deviations. The normalized variable is the state-level rainfall value minus the state-level long-run mean, divided by the state-level standard deviation over 1949-2005.

[21]http://csc.noaa.gov/

[22]http://justiceinmexico.org/

[23]www.seguridadpublicaenmexico.org.mx

possible inconsistencies between their intentions and their subsequent behavior[24].

4.2 Results

In order to make this section more easy to read, the main results are first briefly summarized. The section is organized as follows:

First, intentions are focused on, since their informational content condition the validity of the theoretical interpretation developed above. Results are comforting though since I find a positive and robust correlation between intention to migrate and migration, whatever the broad destination considered (internal migration or migration to the U.S).

Second, I focus on individuals intending to migrate in 2002 and examine together the three possible explanations, emphasized in the presentation of the model, to the fact that some of them had not (yet) materialized their plans in 2005: unexpected shocks, misevaluated costs or liquidity constraints. The first two reasons are found to matter, whereas the last hypothesis is not confirmed by the data. However, the effect of any shock on migration conditional on intention to migrate is smaller than the negative role played by gender. In this subsection, I also discuss migrants selection depending on their observable characteristics.

Third, as a consequence of the latter result, I consider different interpretations of such a gender effect, and present suggestive evidence of costs or constraints specific to women, most plausibly related to their social role as mothers.

Fourth, I discuss migrants' selection as regards their unobserved characteristics, since, as suggested by the model, the misestimation of some of the costs at the intention stage could results in a shift in selection on unobservables at the migration stage. My results, in line with previous studies on the same country, indeed suggest that migrants are positively selected, whereas candidates to emigration seem to be slightly negatively selected.

[24]The symmetrical case is more problematic: a large share of international migrants had no intention to move three years earlier. Nevertheless, the behavior of this specific category of migrants may not be described as inconsistent since the fact that they do not state any intention to migrate may be caused by the planning of the survey. Indeed, at the time of the first wave of the survey, they may still have been gathering information on their income prospects, here and abroad. The mere fact that intentions were not recorded in time does not mean that they had not existed.

4.2.1 Intentions and selection at the intention stage

The first concern raised by the MxFLS data on intentions to migrate, is that given how vague the question is, answers could merely be driven by interviewees' dreams rather than be the result of a rational calculation. In that case, if the question had been perceived as being similar to "would you like to be rich", intentions would certainly be overstated. On the contrary, the data do not support this interpretation: 14.7% of individuals in my sample had the intention to migrate, whatever the destination, and only 2.9% specifically intended to migrate abroad (the U.S. for the overwhelming majority).

More generally, as mentioned above, intentions are mistrusted because they are subjective and do not commit those who state them to anything. Carrying this criticism to an extreme degree, one would claim that intentions would be of no more use than if they were random. Again, no such evidence is provided by the data: First, as is shown in table 1, the correlation between intention to migrate and migration is positive and significant for both destinations (within Mexico and abroad), which would not be the case if interviewees had answered randomly.

Table 1: Probit regressions of migration between 2002 and 2005 on intention to move in 2002

	(1) Migration abroad (2005)	(2) Migration abroad (2005)	(3) Migration within Mexico (2005)	(4) Migration within Mexico (2005)
Intention to migrate abroad (2002) (d)	0.871***	0.708***		
	(0.104)	(0.107)		
Intention to migrate within Mexico (2002) (d)			0.519***	0.453***
			(0.070)	(0.073)
Controls		YES		YES
Constant	-1.872***	-0.809***	-1.775***	-0.836***
	(0.031)	(0.298)	(0.031)	(0.251)
Observations	13038	13036	13038	13036
Pseudo R^2	0.026	0.105	0.024	0.055

Robust standard errors in parentheses (clustered by household)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
(d) dummy variables
Controls included are gender, age, education, regional and geographic dummies

In addition, intentions are found to be correlated with human capital and network variables, consistently with the rational interpretation of intentions within the Roy model developed above, as can be seen in table 2 and 3. In both tables a probit for intentions to migrate (respectively abroad and within Mexico) is run on the whole sample, and then separately for men and women

(column 2 and 3). The fourth column provides the significance level of the difference between the coefficients for men and women. Men have a higher probability to intend to migrate abroad, which justifies the need for running separate regressions for men and women. In fact, male and female intentions, as regards international migration, differ mainly as concerns human capital variables (education). Whereas all three coefficients on education dummies are positive for women (though not significant on the sample restricted to non attritors[25]), they are negative for men. Men with tertiary or vocational education are found to be less likely to have the intention to migrate abroad. This result of opposite selection at the intention stage for men and women relative to education is consistent with previous findings by Kanaiaupuni (2000) studying actual migration.

No such gendered pattern appears when considering intentions to move within Mexico. Intentions are indeed positively and very significantly correlated with education for both men and women[26].

Moreover, as appears in table 2 and 3, intentions are positively correlated with destination specific networks variables, which supports the assumption that networks play a part at the intention stage by supplying individuals with information on the foreign labor market and helping them to form their anticipations.

As noted above, there is no consensus on the nature of selection of Mexican migrants with regard to education. Notwithstanding, some elements of comparison can be emphasized. My results for men are in line with the interpretation proposed by Borjas (1987), using a Roy model with constant migration costs, if one consider that the distribution of wages is less unequal in the U.S. than in Mexico. In that case, indeed, migration gains are expected to be relatively larger for Mexicans at the lowest part of the wage distribution in their home country. The negative selection into migration (at the intention stage, at least), could also reveal a low transferability of skills and diploma between both countries.

[25]When attritors whose intentions are non missing are included in the sample, signs of the coefficients on all education dummies are not modified, but the coefficient on the tertiary education dummy is significant at the 5% level for women, whereas only the coefficient on the vocational education dummy is significant at the 10% level for men, see below for a complete analysis of attrition

[26]The same result is obtained when attritors are included in the regression sample.

Table 2: Probit regressions of intention to migrate abroad in 2002, comparison between men and women

Dependent variable, column (1) to (3): Intention to migrate abroad	(1) All	(2) Men	(3) Women	Diff (2) (3)
Male (d)	0.251***			
	(0.064)			
Age	−0.027	−0.067**	0.023	**
	(0.021)	(0.029)	(0.030)	
Age squared	0.000	0.001	−0.001	**
	(0.000)	(0.000)	(0.000)	
Education: Secondary (d)	0.016	−0.059	0.081	
	(0.087)	(0.130)	(0.108)	
Education: Preparatoria or tertiary (d)	−0.050	−0.297*	0.191	**
	(0.101)	(0.154)	(0.126)	
Education: professional (d)	−0.075	−0.323*	0.179	*
	(0.116)	(0.190)	(0.152)	
Married (d)	−0.257***	−0.267	−0.264**	
	(0.088)	(0.165)	(0.106)	
Household size	0.003	−0.002	0.005	
	(0.016)	(0.022)	(0.022)	
Children < 15 (d)	0.134	0.218	0.077	
	(0.103)	(0.171)	(0.123)	
Elderlies in the household (d)	−0.059	−0.254*	0.076	*
	(0.096)	(0.142)	(0.126)	
Log per capita total expenditures	0.200***	0.207***	0.198***	
	(0.061)	(0.068)	(0.070)	
Locality migration network abroad	0.037***	0.040**	0.038**	
	(0.013)	(0.020)	(0.015)	
Relative(s) in the U.S. (d)	0.572***	0.504***	0.624***	
	(0.079)	(0.104)	(0.097)	
Shocks (1998-2001)	YES	YES	YES	
Geographic controls	YES	YES	YES	
Constant	−3.345***	−2.295***	−4.314***	**
	(0.642)	(0.812)	(0.737)	
Observations	12,879	5,261	7,618	12,879
Pseudo R^2	0.133	0.136	0.147	

Robust standard errors in parentheses (clustered by household)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
(d) dummy variables
Geographic controls include regional dummies and dummies for different urban and rural strata
Shocks include rainfall z-score in 2001, the number of storms (1998-2001), and crime evolution at the state level (1998-2001)
The reference category for education variables is no education or primary education
Column (3) provides a test of equality of coefficients between equations (1) and (2) using seemingly unrelated estimation tools

Table 3: Probit regressions of intention to migrate within Mexico in 2002, comparison between men and women

Dependent variable, column (1) to (3): Intention to migrate in Mexico	(1) All	(2) Men	(3) Women	Diff (2) (3)
Male (d)	0.002			
	(0.041)			
Age	−0.002	0.019	−0.020	
	(0.013)	(0.019)	(0.017)	
Age squared	−0.000	−0.000	0.000	
	(0.000)	(0.000)	(0.000)	
Education: Secondary (d)	0.187***	0.142	0.213**	
	(0.063)	(0.087)	(0.083)	
Education: *Preparatoria* or tertiary (d)	0.369***	0.316***	0.400***	
	(0.075)	(0.101)	(0.102)	
Education: professional (d)	0.654***	0.572***	0.715***	
	(0.079)	(0.108)	(0.107)	
Married (d)	−0.129*	−0.047	−0.162*	
	(0.071)	(0.125)	(0.084)	
Household size	−0.011	0.010	−0.025*	*
	(0.012)	(0.017)	(0.015)	
Children < 15 (d)	−0.009	−0.037	0.000	
	(0.065)	(0.106)	(0.080)	
Elderlies in the household (d)	−0.138**	−0.102	−0.167**	
	(0.066)	(0.098)	(0.083)	
Log per capita total expenditures	0.063**	0.042	0.080**	
	(0.031)	(0.041)	(0.038)	
Locality migration network in Mexico	0.012***	0.006	0.016***	
	(0.003)	(0.005)	(0.004)	
Shocks (1998-2001)	YES	YES	YES	
Geographic controls	YES	YES	YES	
Constant	−1.633***	−1.726***	−1.530***	-1.210***
	(0.385)	(0.507)	(0.482)	(0.019)
Observations	12,879	5,261	7,618	12,879
Pseudo R^2	0.075	0.069	0.086	

Robust standard errors in parentheses (clustered by household)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
(d) dummy variables
Geographic controls include regional dummies and dummies for different urban and rural strata
Shocks include rainfall z-score in 2001, the number of storms (1998-2001), and crime evolution at the state level (1998-2001)
The reference category for education variables is no education or primary education
Column (3) provides a test of equality of coefficients between equations (1) and (2) using seemingly unrelated estimation tools

4.2.2 Who migrates and who does not? Shocks, costs or liquidity constraints?

Since intentions are found to have an unquestionable informational content, the next step consists in understanding the heterogeneity in individual migration behaviors among individuals with intention to migrate in 2002. The theoretical part of this study emphasized three major sets of reasons. First, unexpected and unpredictable shocks could occur before migration plans were carried out and modify individual calculations so that they would cancel or postpone their decision to migrate. Second, costs and constraints to migration could have been misidentified or miscalculated at the intention stage. Third, liquidity constraints may force poorer migrants to postpone their plans.

In order to assess the validity of each interpretation I estimate a bivariate probit model for intentions and migration, and present in tables 4 (for international migration) and 5 (for internal migration) the marginal effects of the right-hand side variables on migration, conditional on the initial intention to migrate[27]. The set of regressors in both equations includes gender, human capital variables (measured in 2002) and geographical controls. Various sets of shocks variables are added to the migration equation, in columns 3 to 12 of each table. Results in tables 4 to 7 read in percentage points. In order to control for potential liquidity constraints, I include a variable for household per capita consumption, proxying for household wealth, geographic controls that proxy for the accessibility to credit facilities, and network variables, that could capture, among other aspects, the possibility for would-be emigrants to fund their travel, as assumed by Borger (2011). The liquidity constraints assumption is specifically discussed (and rejected) in subsection 4.2.3, based on results shown in table 8.

Models presented in columns 3 to 6, in table 4 and 5 include state-level shocks variables. The construction of rainfall z-score has been explained above. The hurricane dummy equals one if the state has been hit by a hurricane between 2002 and 2004. This variable, as well as the number of storms (including hurricanes, but also lower intensity tropical storms) have been constructed using information gathered from the Coastal Services Center database of the National Oceanic and Atmospheric Administration. They thus differ from the earthquake and hurricane dummies included in the models whose results are presented in columns 7 and 8, coming from the MxFLS' community questionnaire. The rationale behind the use of different level data for the same

[27] With above notations marginal effects presented in tables 4 to 8 are of the type $\frac{\partial E[y_2|y_1=1,x]}{\partial x}$.

phenomenon is their complementarity. State level shocks may be interpreted as proxies for actual shocks (a hurricane recorder at the state level increases the probability for a given individual or household to have been directly hit), or proxies for changes in local labor market conditions (and thus capture an indirect effect of adverse economic shocks). Note that, reassuringly enough, the correlation is positive and reasonably large between hurricane variables at the state and community level (0.29)[28].

The nature of shocks included in the next three models (columns 6 to 9) is very different. In an attempt to identify potential exogenous unexpected shocks likely to modify individual's expectations between 2002 and 2005, I used variables related to the evolution of violence and crime within this span of time. Violence is indeed a national issue, with regional peculiarities and an increasing drug-related criminality. The first variable represents the evolution, in percentage, of the number of registered crimes (all included) at the state level. The variation between the 16 states represented in the survey is large (see summary statistics in Appendix), whereas the country average (for all 32 entities) for the same period is characterized by a small decrease (-5%). Variables included in columns 10 and 11 are based on the community informant's subjective assessment of the evolution of violence in the community in the last 12 months (community questionnaire). Finally the last model includes shocks at the household level: the death or illness of a member of the household, and a dummy that equals one for a range of economic shocks (unemployment) and natural disasters.

[28] 17% of communities declared to have been affected by a hurricane in the last three years. In 62% of cases, no hurricane was recorder at the state level (based on an exhaustive list): either hurricanes declared at the community level were only storms (of lower intensity), or recall errors caused hurricanes prior to 2002 to be declared. Note also that hurricanes in 2005 at the state level were not included for lack of the precise survey date, since there is no possibility to control that they occurred before the second wave of the survey. Even without errors in hurricane statements at the community level, this would explain the fact that the correlation is only partial.

Table 4: Marginal effects on the probability to migrate abroad between 2002 and 2005 conditional on intention to migrate abroad in 2002

Marginal effects on Prob(migration abroad= 1|intention to migrate abroad= 1) after bivariate probit	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female (d)	−0.071***	−0.066***	−0.068***	−0.066***	−0.066***	−0.067***	−0.065***	−0.065***	−0.066***	−0.066***	−0.066***	−0.064***
	(0.015)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.013)
State level weather shocks												
Rainfall z-score 2003			−0.012**									
			(0.006)									
Rainfall z-score 2004			−0.020***									
			(0.006)									
High rainfall 2003-04 (d)				−0.004								
				(0.018)								
Hurricane 2002-04 (d)					0.028							
					(0.021)							
Number of storms 2002-04						0.023***						
						(0.009)						
Community level shocks												
Earthquake (d)							−0.042*					
							(0.023)					
Hurricane (d)								−0.035*				
								(0.019)				
Crime and violence												
Crime evolution (state level) 2002-05									−0.000***			
									(0.000)			
Crime increased (community) 2004-05 (d)										−0.003		
										(0.016)		
Violence to women increased (community) 2004-05 (d)											−0.019	
											(0.015)	
Household shocks												
Death/illness 2003-2005 (d)												−0.026**
												(0.013)
Adverse economic shock 2003-2005 (d)												0.009
												(0.018)
Controls	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Estimated probability Prob(migrant 2005= 1|intention 2005= 1)	0.148	0.107	0.106	0.107	0.108	0.108	0.108	0.106	0.107	0.107	0.106	0.079
Observations	13,038	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	12,842

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included are age, education, regional and geographic dummies

Table 5: Marginal effects on the probability to migrate in Mexico between 2002 and 2005 conditional on intention to migrate in Mexico in 2002

Marginal effects on Prob(migration in Mexico = 1|intention to migrate in Mexico = 1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female (d)	−0.022**	−0.018**	−0.019**	−0.018**	−0.018**	−0.019**	−0.018**	−0.019**	−0.018**	−0.018**	−0.018**	−0.023***
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
State level weather shocks												
Rainfall z-score 2003			0.006									
			(0.005)									
Rainfall z-score 2004			−0.009*									
			(0.005)									
High rainfall 2003-04 (d)				−0.007								
				(0.015)								
Hurricane 2002-04 (d)					0.004							
					(0.013)							
Number of storms 2002-04						0.009						
						(0.005)						
Community level shocks												
Earthquake (d)							0.003					
							(0.018)					
Hurricane (d)								0.031**				
								(0.014)				
Crime and violence												
Crime evolution (state level) 2002-05									−0.001***			
									(0.000)			
Crime increased (community)										0.010		
past 12 months (d)										(0.013)		
Violence to women increased											−0.004	
(community) past 12 months (d)											(0.012)	
Household shocks												
Death/illness 2003-2005 (d)												0.021
												(0.014)
Adverse economic shock 2003-2005 (d)												0.029
												(0.020)
Controls	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Estimated probability												
Prob(migrant 2005= 1|intention 2005= 1)	0.103	0.089	0.089	0.089	0.089	0.090	0.089	0.089	0.089	0.089	0.089	0.078
Observations	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	12,842

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included are age, education, regional and geographic dummies

First, as regards weather shocks measured at the state level, I find a negative correlation between normalized precipitations in 2003 and 2004 and international migration, conditional on intentions. These results are consistent with those of Pugatch and Yang (2011) who indeed find that rainfall in Mexico is negatively correlated with Mexican immigration (measured from U.S. data sources as the share of male Mexicans of the U.S. labor force). The interpretation that they propose is twofold:

> We are unable to distinguish whether this effect operates primarily by driving new emigrants out of drought-afflicted Mexican states, or by discouraging return migration from the U.S. to these drought-affected areas.

I am here able to bring pieces of evidence that support the first of these two interpretations, and suggest a third one. Measuring here migration at the source, using origin country data, I am not concerned with return migration issues and observe directly a negative correlation between net migration outflows and rainfall. Nonetheless, the drought interpretation is challenged by the fact that I am presenting marginal effects, conditional on previous intentions to move. Low rainfall, and thus drought could indeed result in increased vulnerability and be an incentive to carry out migration plans earlier. But heavy rainfall, by improving local economic prospects, could make migration plans less or no more profitable by reversing the sign of the expected income differential between the U.S. and Mexico. The fact that 2003 and 2004 happen to have been two especially rainy years supports this latter interpretation. Indeed, z-scores are negative in one state only in 2003 (Sonora) and two states in 2004 (Sinaloa and Morelos). For lack of inter-state variability, however, this interpretation cannot be convincingly tested by splitting the z-scores variables into negative and positive ones. On the other hand, the fact that the number of storms at the state level has a positive and significant impact may speak in favor of an interpretation in terms of increased vulnerability.

At the community level results are different: adverse shocks (earthquake or hurricanes) are associated with a lower propensity to migrate. A possible explanation for observing opposite effects of similar shocks (storms and hurricanes[29]) measured at different levels is the following: at the state level, shocks variables capture indirect effect of weather shocks, and may thus proxy for unexpected adverse changes on state level labor markets or commodity prices. Thus, having

[29]The coefficient on the hurricane dummy is indeed positive and significant for men, whereas hurricanes at the state level do not affect the materialization of women's migration plans, as is shown below, in table 6

less job opportunities locally, individuals (men especially) would be led to migrate earlier. At the community level, on the other hand, survey based variables allow to measure direct effects of natural disasters: it is very likely that individuals living in communities actually hit by a hurricane incurred financial losses which would negatively affect their ability to migrate through costs effects (no money left to pay for the travel). Note that in the internal migration case, on the contrary, the consequence of being directly hit by a hurricane is to increase the probability to move conditional on intention. This finding is consistent with the cost interpretation, since migration costs are undoubtedly much higher in the international migration case.

As for crime variables, they are not found to affect migrations plans, neither at the state-level (objective figures), nor at the community level (subjective assessments). The significance of the coefficient in column 9 is exclusively driven by the outburst of crime in the Sonora state, with a 146% increase between 2002 and 2005 (which is actually a direct consequence of anti-immigration policies in the U.S. at the same time).

Then, among household shocks, death or illness of one household member is found to affect negatively the probability to comply with one's intention to move, or delay migration plans[30].

In conclusion, shocks are found to affect materialization of migration plans abroad, but in general do not alter internal migration plans. This result is not surprising, internal migration being characterized by lower costs and benefits and in many cases being in all probability driven by non economic motives (for example marriage, family, health).

In order to check the consistency of my findings for shocks, I computed their marginal effects on the probability to migrate conditional on *not* having the intention to migrate in 2002 (either within Mexico or abroad). Results for international migration are presented in Appendix, table 16 and prove consistent with the above findings: shocks are found to affect those who had no intention to move in a similar way (same signs on all coefficients) as those who had the intention to migrate, but the magnitude of the coefficient is unsuprisingly much lower.

[30]Unfortunately, the data do not allow to know whether this correlation could be explained by the fact that illness of the individual with the intention to migrate would force him to change his plans since the affected household member cannot be precisely identified. A solution would consist in considering only death shocks, which could only affect another household member, but they concern such a low percentage of individuals (5.3%) that they cannot be used separately.

Shocks and education Shocks are not expected to have homogeneous impacts on migration plans, depending on individual observable characteristics, and in particular education. Education may indeed be considered as a proxy for wealth, less endogenous than income or consumption as regards migration propensity. Consequently, individuals with higher education levels might be less affected by adverse shocks than those with no education.

This is indeed what is found when interacting shocks variables with an education dummy being equal to one for secondary or tertiary education. Table 17 in Appendix, shows marginal effects on the probability to migrate abroad conditional on intention to do so, obtained after bivariate probit regressions, with interaction terms between weather shocks and shocks at the household level and education[31]. Education is not found to affect directly the materialization of migration plans, which suggests that the component explained by education of the cost and both wage functions may be accurately estimated by individuals at the intention level. Besides, education, very likely because it proxies for wealth, seems to attenuate or cancel the impact of most weather shocks: the marginal effects of the interaction term and the shocks variables are of opposite sign and comparable magnitude for rainfall in 2004, the hurricane and storms variables at the state and community level (though the interaction term is estimated with much imprecision for the latter). On the other hand, rainfall in 2003 and death or illness shocks at the household level are found to have similar impacts on individuals' plans whatever their education level.

Shocks and gender Besides the influence of shocks on migration, the most striking result in both tables 4 and 5 is the negative and significant coefficient on the female dummy: being a woman reduces by almost one half (around 7 percentage points) the probability to migrate abroad between 2002 and 2005, and one fifth (2 percentage points) the probability to migrate in Mexico, conditional on having the intention to migrate in 2002. Coefficients on the female dummy are robust to the inclusion of any set of shocks in the model. Moreover, being a woman cuts by two thirds the probability to move abroad, which is more than the larger marginal effect of any shock (40% for the earthquake dummy at the community level). Being a women is a greater handicap to be able to carry out one's migration plans within a three year span than

[31]Note that the education dummy could not be interacted with the variable for earthquake at the community level because of a too small number of observations.

suffering an earthquake.

Consistently with an interpretation in terms of costs (the constraints and migration costs linked to being a woman would be underestimated at the intention stage), the marginal effect of the female dummy is smaller in the internal migration case, but it is still negative and robust to the introduction of shocks.

A natural extension thus consists in adding interactions between shocks variables and the female dummy variable: results are presented in table 6 and 7 for international and internal migration. Models including different sets of crime shocks are not shown here since none of these variables are found to significantly affect migration.

Consider first international migration (table 6): with the exception of the first model (including rainfall z-scores), all results suggest that shocks do not affect more women than men. In models 2 and 4, the interaction term is not significant, shocks are thus found to affect similarly men and women. In models 3 and 5 to 7, the interaction term and main effect are of opposite signs and comparable magnitude, suggesting that shocks affect exclusively men's migration (conditional on intention), although the interaction term may not be significant due to the relatively small number of observations. In all those specifications, the magnitude and significance of the female dummy remains constant.

The first specification is worth going back over: once adding an interaction term between the female dummy and rainfall in 2004, the female main coefficient is not significant, whereas the interaction term suggests that women's migration plans are twice more affected than men's by rainfall. Nonetheless, this specification is problematic because of the high and positive correlation (0.53)[32] between precipitations in 2004 and the center region dummy. Actually, the cumulative negative effect of rainfall on female migration seems to be driven by the exceptionally heavy rainfall that affected the two states of Mexico and Morelos, as well as the Federal District[33]. A plausible interpretation is thus that women living in the center region (and in particular in Mexico City) with emigration plans are found to be less likely to have them materialized within three years, may be because of a larger set of local job opportunities than in other regions.

[32]For that reason, the set of controls in this specification (in all tables) does not include the center region dummy.

[33]Results are indeed similar, except that the marginal effect of the female dummy is significant, when the center region dummy instead of the 2004 rainfall variable is interacted with the female dummy.

Note that adding interaction terms reveals the positive marginal effect of the hurricane variable at the state level on the conditional probability to migrate for men, whereas it does not affect women's migration.

Table 7 presents the results of the same regressions on the probability of internal migration, conditional on intention to move in Mexico. Similar remarks can be made for the hurricane dummy at the state level. The main marginal effect on the female dummy is significant in the last three specifications only, and none of the marginal effect of shocks variables and interaction terms, except the hurricane and 2004 rainfall variables are significant.

4.2.3 Lower materialization of women's migration plans

Table 6: Marginal effects on the probability to migrate abroad between 2002 and 2005 conditional on intention to migrate abroad in 2002

Marginal effects on Prob(migration abroad = 1|intention to migrate abroad = 1) after bivariate probit	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female (d)	−0.037	−0.070***	−0.058***	−0.068***	−0.072***	−0.067***	−0.068***
	(0.023)	(0.018)	(0.013)	(0.018)	(0.014)	(0.014)	(0.015)
State level weather shocks							
Rainfall z-score 2003	−0.013*						
	(0.008)						
Female*rainfall z-score 2003	−0.000						
	(0.008)						
Rainfall z-score 2004	−0.013*						
	(0.007)						
Female*rainfall z-score 2004	−0.018**						
	(0.008)						
High rainfall 2003-04 (d)		−0.006					
		(0.021)					
Female*high rainfall 2003-04 (d)		0.004					
		(0.020)					
Hurricane 2002-04 (d)			0.061*				
			(0.033)				
Female*hurricane 2002-04 (d)			−0.054**				
			(0.021)				
Number of storms 2002-04				0.023*			
				(0.012)			
Female*number of storms 2002-04				0.001			
				(0.014)			
Community level shocks							
Earthquake (d)					−0.066**		
					(0.023)		
Female*earthquake (d)					0.075		
					(0.081)		
Hurricane (d)						−0.043**	
						(0.022)	
Female*hurricane (d)						0.021	
						(0.036)	
Household shocks							
Death/illness 2003-2005 (d)							−0.041***
							(0.015)
Female*death/illness 2003-2005 (d)							0.049
							(0.034)
Adverse economic shock 2003-2005 (d)							0.014
							(0.024)
Female*economic shock 2003-2005 (d)							−0.012
							(0.024)
Controls	YES	YES	YES	YES	YES	YES	YES
Estimated probability							
Prob(migrant 2005= 1|intention 2005= 1)	0.106	0.107	0.108	0.108	0.108	0.106	0.079
Observations	13,036	13,036	13,036	13,036	13,036	13,036	12,842

Marginal effects; Standard errors in parentheses ; (d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls included are age, education, regional and geographic dummies

Table 7: Marginal effects on the probability to migrate in Mexico between 2002 and 2005 conditional on intention to migrate in Mexico in 2002

Marginal effects on Prob(migration in Mexico = 1|intention to migrate in Mexico = 1) after bivariate probit	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female (d)	−0.032	−0.037	−0.014	−0.012	−0.015*	−0.019*	−0.021**
	(0.022)	(0.018)	(0.009)	(0.013)	(0.009)	(0.010)	(0.009)
State level weather shocks							
Rainfall z-score 2003	0.005						
	(0.007)						
Female*rainfall z-score 2003	0.002						
	(0.007)						
Rainfall z-score 2004	−0.012**						
	(0.006)						
Female*rainfall z-score 2004	0.005						
	(0.006)						
High rainfall 2003-04 (d)		−0.019					
		(0.021)					
Female*high rainfall 2003-04 (d)		0.023					
		(0.019)					
Hurricane 2002-04 (d)			0.025				
			(0.020)				
Female*hurricane 2002-04 (d)			−0.032**				
			(0.016)				
Number of storms 2002-04				0.013			
				(0.008)			
Female*number of storms 2002-04				−0.007			
				(0.009)			
Community level shocks							
Earthquake (d)					0.017		
					(0.029)		
Female*earthquake (d)					−0.023		
					(0.027)		
Hurricane (d)						0.028	
						(0.018)	
Female*hurricane (d)						0.004	
						(0.019)	
Household shocks							
Death/illness 2003-2005 (d)							0.032
							(0.021)
Female*death/illness 2003-2005 (d)							−0.016
							(0.018)
Adverse economic shock 2003-2005 (d)							0.028
							(0.027)
Female*economic shock 2003-2005 (d)							0.002
							(0.026)
Controls	YES	YES	YES	YES	YES	YES	YES
Estimated probability							
Prob(migrant 2005= 1|intention 2005= 1)	0.089	0.089	0.090	0.090	0.090	0.089	0.077
Observations	13,036	13,036	13,036	13,036	13,036	13,036	12,842

Marginal effects; Standard errors in parentheses ; (d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls included are age, education, regional and geographic dummies

Table 8: Marginal effects on the probability to migrate abroad between 2002 and 2005 conditional on intention to migrate abroad in 2002, for men and women

Marginal effects on Prob(migration abroad= 1|intention to migrate abroad= 1) after bivariate probit	(1) All		(2) Men		(3) Women	
	mfx	se	mfx	se	mfx	se
Female (d)	-0.060^{***}	(0.013)				
Age	-0.002	(0.002)	0.000	(0.004)	-0.002	(0.002)
Age squared	-0.000	(0.000)	-0.000	(0.000)	0.000	0.000
Education: Secondary (d)	-0.005	(0.010)	-0.011	(0.021)	-0.002	(0.007)
Education: *Preparatoria* or tertiary (d)	0.027	(0.017)	0.049	(0.035)	0.011	(0.011)
Education: professional (d)	-0.017	(0.015)	-0.033	(0.031)	-0.005	(0.009)
Married (d)	0.009	(0.013)	0.001	(0.029)	0.004	(0.007)
Household size	0.009^{***}	(.003)	0.013^{***}	(0.005)	0.005^{**}	(0.002)
Children < 15 (d)	-0.042^{***}	(0.016)	-0.028	(0.028)	-0.029^{**}	(0.014)
Elderlies in the household (d)	-0.020^{**}	(0.010)	-0.063^{***}	(0.021)	-0.001	(0.006)
Locality migration network abroad	0.005^{***}	(0.002)	0.007^{**}	(0.003)	0.002^{*}	(0.001)
Relative(s) in the U.S. (d)	0.034^{***}	(0.010)	0.018	(0.019)	0.034^{***}	(0.012)
Log per capita total expenditures	0.005	(0.005)	0.005	(0.011)	0.003	(0.003)
Geographic controls	YES		YES		YES	
Shocks between 2002 and 2005	YES		YES		YES	
Estimated probability						
Pr(intention 2002 = 1| non migrant 2005= 1)	0.078		0.124		0.037	
Observations	12,689		5,185		7,504	

Robust standard errors in parentheses (clustered by household)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
(d) dummy variables
The reference category is no education or primary education for education variables
Shocks: Rainfall z-scores 2003 and 2004, number of storms, earthquake (community), death or illness (household)
Geographic controls: historic region dummy and rural/urban strata dummies

In table 8, I present results from the same model (bivariate probit) as in table 4 for international migration, with additional controls for demographic characteristics of the household in 2002. Marginal effects read the same as in previous tables (percentage points). A set of shocks variables is still controlled (see the list at the bottom of the table) for but marginal effects are not shown, and separate regressions are run for men and women.

First note that the variable proxying for household wealth has no effect on the probability to migrate conditional on intentions. This suggests that liquidity constraints may not significantly matter to explain the materialization of migration plans.

Second, one of the main findings is that having children under 15 is negatively associated with women's migration (the marginal effect is negative but not significant for men). This result supports the hypotheses made by Kanaiaupuni (2000) relative to the social and financial

constraints to the emigration of mothers (quoted in section 2). It is also consistent with the earlier results of Donato (1993) who finds a negative quadratic relationship between age and the probability for women to migrate. A similar pattern for age is also found here (not shown), but is entirely absorbed by the children dummy[34]. On the other hand, having relatives in the U.S. is positively and significantly correlated with women's migration only. These results are consistent with the interpretation that women incur costs and constraints limiting their ability to migrate, that are partly misestimated at the intention stage. In particular women would not take totally into account the constraints caused by having children in their migration plans (at the intention stage). Moreover, the fact that women seem to rely more than men on family networks to migrate abroad is consistent with the family reunification motive for female migration, making women "associational" migrants ((Kanaiaupuni, 2000) quoting a term coined by (Balan, 1981)).

Of course, alternative explanations could be accepted: first, women could have different preferences concerning the time of their migration. Since no time limit is explicitly included in the intention question, one could imagine that women systematically intend to migrate in a further future than men, in particular when children would have grown up.

Nonetheless, if different preferences were the whole story, since we are observing migration flows, and under the assumption that preferences are relatively stable over time, we should not observe any difference in the size of male and female migrant cohorts between 2002 and 2005 (conditional on intention). The only difference being that female migrants would have waited longer from the time they first intended to migrate. It is however not possible to definitely reject this interpretation, if changes in preferences over time were to result in lower intention to migrate for older women.

Another indication is provided by the age distribution of men and women with intention to migrate, compared to the one of male and female migrants (see figure 2 in Appendix). Indeed, the share of individuals with intention to migrate abroad aged between 25 and 35 years is larger for women than for men. The highest density is around 20 years old for men, and it then steeply declines. This finding is consistent with both interpretations, in terms of constraints delaying or preventing women's migration, or different time preferences: women seem to be willing to migrate as young as men, but they do not migrate as quickly as them, thus swelling the density for ages up to 35 years. However, the same density plotted for migration abroad seems to rather

[34]Donato (1993) controls for the presence of children in the household but not for own children

support the first interpretation: if preferences were to explain the difference between male and female age distributions at the intention stage, we would expect the mode of the age distribution for women at the migration stage to be shifted to the right, as compared to men's. And yet, modes for male and female age distributions coincide at the migration stage, around 23 years old. Figure 3 shows that age distributions for intentions are not much affected when including attritors in the sample.

Another element supports the interpretation in terms of miscalculated costs and constraints: if women had longer term prospects and stated in 2002 intentions to migrate in a more remote future than men, they would then be expected to keep on having the intention to move more longer than men, and in particular three years later.

Table 9: Marginal effects on the probability to have the intention to migrate in 2005 conditional on not having migrated between 2002 and 2005 : subsample of individuals with intention to migrate in 2002

Marginal effects on Prob(intention 2005= 1|non migrant 2005= 1) obtained from a bivariate probit	Marginal effects	Standard errors
Female (d)	−0.026	(0.031)
Age	0.000	(0.010)
Age squared	−0.000	(0.000)
Education 2ry and higher (d)	0.075**	(0.034)
Married (d)	−0.032	(0.047)
Household size	−0.011	(0.009)
Children < 15 (d)	−0.048	(0.043)
High rainfall 2003-2004 (d)	0.077**	(0.033)
Human shock 2003-2005 (d)	0.017	(0.048)
Economic shock 2003-2005 (d)	0.145**	(0.063)
Geographic controls	YES	
Observations	1670	

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Geographic controls: historic region dummy and rural/urban strata dummies

Results shown in table 9 are marginal effects on the probability to have the intention to migrate in 2005 conditional on not having migrated between 2002 and 2005, obtained after a bivariate probit for migration between 2002 and 2005 and intentions to migrate in 2005 on the subsample of those who had the intention to migrate in 2002. Due to the small number of observations, the destination (internal vs international migration) is not taken into account (as in tables 10 and 11). The female dummy is not significant: women do not persist significantly

longer than men in their migration plans, which suggests that preferences regarding the timing of migration may not be so much different from men and women

It could also be argued that a household model would be better suited to the analysis of Mexican migration. Women's intentions to migrate would be viewed as depending on a decision taken at the household level. Indeed, as appears in table 10, intentions to migrate are correlated within households[35]. Note however that this correlation exists for both men and women: the interaction term is not significant, women's intentions to migrate are not found to depend more than men's on a spouse or parent's own migration plans.

Table 10: Correlation between intention to move of individuals of different sex within the same household

Probit regression, individual level Dependent variable	Intention to migrate, 2002
Intention to move of another household member, different sex	0.681^{***} (0.073)
Female (d)	-0.044 (0.050)
Intention to move of another household member, different sex* Female	0.002 (0.047)
Controls	YES
Constant	-0.916^{***} (0.172)
Observations	13185

Robust standard errors in parentheses (clustered by household)
(d) dummy variables
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included are age, education, regional and geographic dummies

A slightly different explanation could be that women would voice more often than men a household decision to migrate. In order to test for this interpretation, I investigate whether women's intentions to migrate are associated with a higher probability to observe male migration in the same household. As presented in table 11, this seems to be the case, but as in the previous table, these results support more a joint decision model for migration than a specific meaning to be attributed to female intentions. Indeed, the lower part of the table presents similar results for the reverse case: female migration is positively correlated with men's intention to move (within

[35]Note that destination could not be taken into account either in this regression, because of the reduced sample size. The regression sample indeed includes only households of size at least equal to 2, including two persons of opposite sex.

the same household). Note that here again, the sample is restricted to individuals living with at least one person of opposite sex whose intentions as regards migration are also recorded in 2002 (1,662 individuals being the only adult of their household, 5,458 women living in households with at least one other adults but no adult male member and 178 men living in households with two or more adults but no adult women are excluded from the regression sample).

Table 11: Correlation between male migration between 2002 and 2005 and female intentions to migrate in 2002 at the household level

Probit regression, household level Dependent variable	At least one male migrant member between 2002 and 2005	
At least one woman with intention to move in 2002	0.215^{**} (0.093)	0.167^{*} (0.090)
Controls	NO	YES
Constant	-1.241^{***} (0.049)	-1.864^{***} (0.275)
Dependent variable	At least one female migrant member between 2002 and 2005	
At least one man with intention to move in 2002	0.256^{***} (0.080)	0.198^{**} (0.081)
Controls	NO	YES
Constant	-0.979^{***} (0.045)	-2.757^{***} (0.289)
Observations	5081	5021

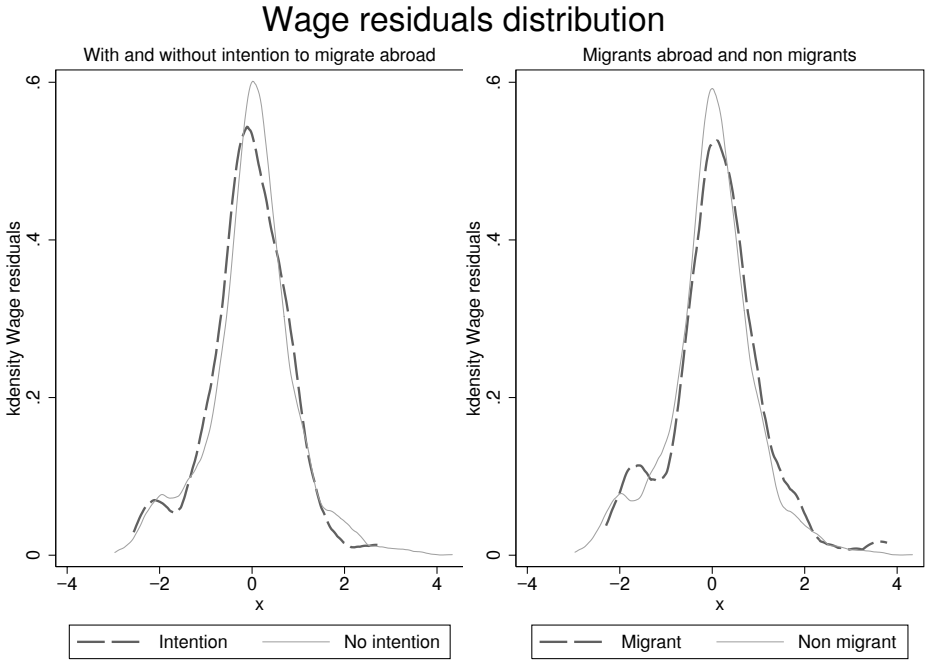
Robust standard errors in parentheses (clustered by locality)
(d) dummy variables
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included are the age of the head, highest level of education of any adult household member weather shocks at the state level, household shocks, regional and geographic dummies

The latter two tables unsurprisingly emphasize the fact that members of the same household coordinate their migration decision. However, this obvious result does not explain why women do migrate less than men conditional on initial plans.

4.2.4 Selection on unobserved characteristics

Figure 1 show the distribution of residual wages for, on the one hand, individuals with intention to move abroad, and intention to stay, and on the other hand, international migrants and non migrants. Whereas the distribution seems to be slightly shifted to the left for those with intention to move (negative selection), the opposite is found for migrants, who thus appear to be positively selected (as compared to non migrants). This result is consistent with Borger

Figure 1: Selection on unobserved characteristics



(2011) and is confirmed by the comparison of the mean of wage residuals reported in table 12, although tests indicate that the difference between any two groups is not significant.

The interpretation of the above finding is further limited since the analysis is relevant only for the individuals who report to work in 2002 (55,9% among those with no intention to move abroad and 61,5% among those with intention to move abroad). In addition missing wage data entail the loss of 13% observations among those who declare to be working. The share of missing data is as high as 20% in the subgroup of individuals who were to migrate abroad between 2002 and 2005. For all these reasons, no definitive statement can be drawn from the above analysis, since it concerns a minority of the overall sample (those who migrate, and with non missing wage data).

We can at most draw from this exercise an indication that migrants who worked in 2002 are endowed with characteristics other than the usual determinants of wages, such as experience, gender, education or region, that therefore are unobserved and explain their higher pre-migration wage, whereas individuals who had the intention to migrate in 2002 and were working at the time seem to be negatively selected with respect to the same unobserved characteristics.

Table 12: Mean wage residuals

	With intention to migrate abroad	With no intention to migrate	International migrant	Non migrant
Mean	−0.070	0.003	0.092	−0.005
Standard error	(0.058)	(0.015)	(0.081)	(0.014)
Observations	224	4001	151	4375

4.2.5 Attrition analysis

Attrition is a major problem when using panel data to address migration issues since migration is very likely the most important cause of attrition. Another concern in my study is the loss of observations due to missing intentions data at the baseline survey. Since these two sources of data loss are driven by very different reasons, I analyse them separately.

The initial sample, in the first survey wave is made of 19,177 adults age 18-64 years. Among them, 15,917 (83%) only have intentions data. For the remaining 3,260, the section on intentions to migrate could not been fulfilled because they were not present in the households at the time of the survey. The problems that may arise from this reduction of my sample of interest are potentially biased estimation results, if these individuals were more likely to both have the intention to migrate and actually migrate. In that case, I would overestimate discrepancies between intentions and subsequent actions, by considering only those who were at home in 2002 and could be interviewed. However this may not be a major issue, since the main objective of this article is to understand the noted discrepancies and not assess their frequency or probability. In this respect I can consider that my sample of interest is made of the 15,917 individuals with recorded answers to the intention question.

Concerning attrition now, the re-contact rate of these 15,917 individuals with non missing intentions data is around 92%. 1,237 individuals are lost. Our 15,917 individuals of interest form 7,474 households in 2002. 9.58% if them (673) are affected by attrition. In 668 cases, the whole household is lost, but they are mostly small households (84% are made of one or two individuals). Not surprisingly the attrition rate is higher in two of the three states at least partly included in the greater Mexico city (Federal district: 16.23%, Puebla: 13.55%). Similarly, the attrition rate is higher in urban areas. There is no gender bias since we lose 8.27% of women and 8.13% of men.

As expected, as regards intentions to migrate, attritors have higher initial intentions to move (24.35% against 16.48% of non attritors).

Then, a third source of data loss is due to the temporary absence of some individuals when surveyors came and visited their household in 2005. They are individuals who are declared to be currently residing household members by other members of the household, or "movers" not considered migrants according to the definition I chose (as stated above, I do not consider migrants individuals who moved within the same locality, for example across the street). In both cases, no information was collected on their past migration status (between 2002 and 2005) so that I cannot identify them as return migrants or "stayers". I thus chose to drop them from my sample. The percentage of these additionally lost individuals with intention to move is higher (22.62%). There is a strong gender bias among these additionally lost observations: men are more often lost (14.31% of them are not in the final sample) than women (6.58%). In the remainder I incorporate them into the category of attritors.

In the results previously shown, I chose to drop all those whose migration status between 2002 and 2005 could not be identified (including individuals who died between the two dates). The major problem is that this obviously leads me to underestimate migration, and may bias my results. To test their robustness I apply three different treatments to attritors and consider that:

1. all attritors migrated within Mexico, which amounts to overestimating internal migration and underestimating international migration

2. all attritors migrated abroad, with symmetrical implications: underestimation of internal migration and (large) overestimation of international migration

3. all attritors who had the intention to migrate abroad actually did it, and that the remaining part of them migrated within Mexico.

All regressions whose results are shown above are thus re-ran on the sample augmented with attritors. Note that the few specifications needing household variables measured in 2005 could not be tested (household shocks, tables 4 to 7, last column).

The reason why I apply the third treatment is the following: since what I am interested in is understanding the discrepancies between intentions to move and actual moves, and particularly

understanding what made individuals with intention to move not to have migrated (yet) three years later, the way attrition could most bias the results would be the case where all attritors with intention to move actually did so. In order to assess the robustness of my result to this potential bias, I re-ran all regressions considering now that all attritors (whose household has not been surveyed in 2005) with intention to move abroad were current migrants abroad in 2005. In doing this I very likely overstate the consistency of individual behaviors with their previous intentions. Nonetheless, results are not challenged by this highly hypothetical assumption aimed at representing the highest bound of the correlation between intentions and actual behaviors[36].

I thus find that results presented in this article are robust to different treatment of attritions, and may not be subject to problematic biases due to attrition.

5 Conclusion

In this article, I try to revisit the applications of the Roy model to migration by considering together subjective data on intentions to migrate and objective migration facts. The comparison between intentions and migration behaviors reveals discrepancies, that can however be explained in a rational setting. Indeed, unexpected shocks or miscalculated costs could cause individuals to modify their initial plans. A simple extension of Borjas' framework shows that omitting non-constant costs in the model may modify and even reverse the sign of migrants' initial self-selection on the basis of individual unobservable characteristics. The empirical analysis based on panel data from the Mexican Family Life Survey first bring elements that support the rational interpretation of migration intentions. Second, I find that migration plans of Mexicans are indeed affected by different shocks, and in particular weather shocks at either state or community level. The estimated impact of weather shocks is consistent with the results presented by Pugatch and Yang (2011), who study Mexican immigration to the U.S. and use the same source of rainfall data together with U.S. labor market data. Nevertheless, the larger marginal effect on the probability to migrate conditional on initial intention is due to gender: women have a much lower propensity to migrate abroad in a three year span, as compared to men, whereas a non negligible part of them contemplates emigrating. This finding suggests

[36]Minor changes are observed, and in particular, including attritors in the estimation sample of the male intention to move abroad equation makes negative coefficient on the dummies for the highest levels of education significant.

that social constraints or specific costs incurred by women may not be fully taken into account at the intention stage, and that gender inequalities affect women's direct access to the benefits from migration: most women who want to move are either constrained to stay or delay the materialization of their migration plans.

A side result of this article is thus that intentions, though fully rational, are not a good proxy for actual migration. In particular, using intention data would lead to a large overestimation of female migration.

As for selection relative to education, results presented in this study suggest interesting gender-specificities: selection occurs at the intention stage, and women are found to positively self-select into international migration, whereas men are negatively selected. As for internal migration, results suggest positive selection at the intention stage for both men and women.

Moreover shocks, costs and constraints are unsurprisingly found to be more decisive for international migration than internal migration. Finally, as concerns selection on unobservables, estimation on wage equations for the subsample of migrants employed in 2002 provide indications of a shift in migrant selection on unobservables between the two stages of migration decision, with a negative selection at the intention stage, and a positive selection of actual migrants. This finding, though not significant, is consistent with previous studies on Mexican migrants.

The above findings may suggest new avenues of research on migration taking more into account original information conveyed by subjective data in general, and migration intentions in particular. At last, the welfare impact as well as the very nature of the probable obstacles to female migration remains to be assessed.

Appendix

Unemployment can be dealt with in the following way: since we know whether individuals have the intentions to move or stay even though they are unemployed, if they are unemployed and still want to stay, the wage they would receive in Mexico would they participate in the local labor market is expected to be above their offered wage in the US. Non participation to the Mexican labor market can be accounted for by introducing a reservation wage w^*. Note that it is assumed that nobody wants to migrate in the purpose of non participating to the U.S. labor market, which implies that the reservation wage is assumed to be common to both labor markets.

A generalization of the model is thus,

$$\begin{cases} \text{Intention to migrate:} & \text{if} \quad w_{\text{US}} > \max(w_{\text{Mex}}, w^*) \\ \text{Intention to stay:} & \text{if} \quad \max(w_{\text{Mex}}, w^*) > w_{\text{US}} \end{cases}$$

Table 13: List of variables

Variable	definition
Education: primary	equal to 1 if the interviewee's highest education is primary schooling; 0 otherwise
Education: secondary	equal to 1 if the interviewee's highest education is secondary schooling; 0 otherwise
Education: Preparatoria or tertiary (d)	equal to 1 if the interviewee's highest education is tertiary schooling; 0 otherwise
Education: vocational (d)	equal to 1 if the interviewee's highest education is vocational; 0 otherwise
Married (d)	equal to 1 if the interviewee is married in 2002; 0 otherwise
Household size	number of persons living in the household in 2002
Children < 15 (d)	equal to 1 if the interviewee has children aged less than 15 years in 2002; 0 otherwise
Elderlies in the household (d)	equal to 1 if at least one household member is aged 65 and over in 2002; 0 otherwise
Log per capita total expenditures	annual amount declared in 2002 for the household
Relative(s) in the U.S. (d)	equal to 1 if the interviewee has relatives in the U.S. in 2002; 0 otherwise
Locality migration network in Mexico	percentage of internal migrants among all adults from other households in the same locality
Locality migration network abroad	percentage of migrants abroad among all adults from other households in the same locality
Regional dummies	equal to 1 if the household lives in 2002 in one of the following states; 0 otherwise:
Border region (d)	Baja California Sur, Coahuila, Nuevo León, Sinaloa and Sonora
Center region (d)	México, Morelos, Oaxaca, Puebla and Distrito Federal
Periphery region (d)	Veracruz and Yucatán
Historic region (d)	Durango, Guanajuato, Jalisco and Michoacán
Urban strata1 (d)	equal to 1 if the household lives in 2002 in a locality with more than 100,000 inhabitants; 0 otherwise
Urban strata2 (d)	equal to 1 if the household lives in 2002 in a locality with 15,000 to 100,000 inhabitants; 0 otherwise
Urban strata3 (d)	equal to 1 if the household lives in 2002 in a locality with 2,500 to 15,000 inhabitants; 0 otherwise
Rural strata (d)	equal to 1 if the household lives in 2002 in a locality with less than 2,500 inhabitants; 0 otherwise
Adverse economic shock (d) 2002-05	equal to 1 if the household suffered unemployment, natural shock, loss of animals or crops between 2002 and 2005; 0 otherwise
Death or illness shock (d) 2002-05	equal to 1 if the household suffered illness or death of any of its members between 2002 and 2005; 0 otherwise
Rainfall z-score 2003	state level normalized rainfall in 2003
Rainfall z-score 2004	state level normalized rainfall in 2004
High rainfall 2003-04(d)	equal to 1 if $\min(\text{rainfall z-score } 2003, \text{rainfall z-score } 2004) > 2$; 0 otherwise
Hurricane 2002-04 (d)	equal to 1 if at least one hurricane hit part of state area between 2002 and 2004; 0 otherwise
Number of storms 2002-04	Number of hurricanes and tropical storms hitting state area between 2002 and 2004; 0 otherwise
Crime evolution 2002-05	State level percentage change in the number of registered crimes between 2002 and 2005
Crime increased past 12 months (d)	Community level, equal to 1 if crime is estimated to have increased in 2005; 0 otherwise
Violence to women increased, past 12 months (d)	Community level, equal to 1 if violence to women is estimated to have increased in 2005, 0 otherwise

Table 14: Summary statistics: individual and household variables

	mean / pct	standard error
Male (d)	40.9	
Age	36.676	13.029
No education (d)	9.1	
Education: Primary (d)	45.8	
Education: Secondary (d)	27.8	
Education: *Preparatoria* or tertiary (d)	11.8	
Education: vocational (d)	5.5	
Married (d)	70.4	
Household size	5.018	2.187
Children < 15 (d)	53.8	
Elderlies in the household (d)	14.2	
Log per capita total expenditures	8.917	0.904
Relative(s) in the U.S. (d)	36.1	
Locality migration network in Mexico	12.689	7.008
Locality migration network abroad	2.354	2.667
Border region (d)	33.1	
Historic region (d)	27.3	
Center region (d)	25.2	
Periphery region (d)	14.4	
Urban strata1 (d)	33.1	
Urban strata2 (d)	9.4	
Urban strata3 (d)	11.4	
Rural strata (d)	46.2	
Human shock 2002-05 (d)	17.3	
Adverse economic shock 2002-05 (d)	11.1	
Observations	$13,038$	

(d) dummy variables

Table 15: Summary statistics: state and community level shock variables

	mean / pct	standard error
Rainfall z-score 2003	2.214	1.454
Rainfall z-score 2004	1.342	1.491
High rainfall 2003-04 (d)	62.5	
Hurricane 2002-04(d)	18.8	
Number of storms 2002-04	0.875	0.957
Crime evolution 2002-05	3.631	41.910
Observations	16	
Earthquake (d)	4.7	
Hurricane (d)	18.0	
Crime increased, past 12 months (d)	37.3	
Violence to women increased, past 12 months (d)	48.7	
Observations	150	

(d) dummy variables

Table 16: Marginal effects on the probability to migrate abroad between 2002 and 2005 conditional on no intention to migrate

Marginal effects on Prob(migration abroad= 1|intention to migrate= 0) after bivariate probit	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female (d)	−0.023***	−0.022***	−0.021***	−0.022***	−0.022***	−0.022***	−0.021***	−0.021***	−0.022***	−0.022***	−0.021***	−0.023***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
State level weather shocks												
Rainfall z-score 2003			−0.003**									
			(0.002)									
Rainfall z-score 2004			−0.004***									
			(0.001)									
High rainfall 2003-04 (d)				−0.000								
				(0.005)								
Hurricane 2002-04 (d)					0.007							
					(0.006)							
Number of storms 2002-04						0.006***						
						(0.002)						
Community level shocks												
Earthquake (d)							−0.010**					
							(0.005)					
Hurricane (d)								−0.010**				
								(0.005)				
Crime and violence												
Crime evolution (state level) 2002-05									−0.000*			
									(0.000)			
Crime increased (community)										−0.001		
2004-05 (d)										(0.004)		
Violence to women increased											−0.005	
(community) 2004-05 (d)											(0.004)	
Household shocks												
Death/illness 2003-2005 (d)												−0.007**
												(0.003)
Adverse economic shock 2003-2005 (d)												0.003
												(0.006)
Controls	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Estimated probability Prob(migrant 2005= 1|intention 2005= 1)	0.028	0.022	0.021	0.022	0.022	0.021	0.021	0.021	0.021	0.22	0.021	0.019
Observations	13,038	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	13,036	12,842

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Controls included are age, education, regional and geographic dummies

Table 17: Marginal effects on the probability to migrate abroad between 2002 and 2005 conditional on intention to migrate abroad in 2002, interactions between shocks and education

Marginal effects on Prob(migration abroad = 1|intention to migrate abroad = 1) after bivariate probit	(1)	(2)	(3)	(4)	(5)	(6)
Education, secondary or tertiary (d)	−0.003	0.014	0.012	0.029	0.004	0.005
	(0.024)	(0.018)	(0.014)	(0.019)	(0.014)	(0.013)
State level weather shocks						
Rainfall z-score 2003	−0.009					
	(0.008)					
Education, secondary or tertiary (d)* rainfall z-score 2003	−0.008					
	(0.010)					
Rainfall z-score 2004	−0.027***					
	(0.007)					
Education, secondary or tertiary (d)* rainfall z-score 2004	0.021**					
	(0.010)					
High rainfall 2003-04 (d)		−0.000				
		(0.020)				
Education, secondary or tertiary (d)* high rainfall 2003-04 (d)		−0.009				
		(0.021)				
Hurricane 2002-04 (d)			0.056*			
			(0.032)			
Education, secondary or tertiary (d)* hurricane 2002-04 (d)			−0.043*			
			(0.025)			
Number of storms 2002-04				0.034***		
				(0.012)		
Education, secondary or tertiary (d)* number of storms 2002-04				−0.025*		
				(0.015)		
Community level shocks						
Hurricane (d)						−0.048**
						(0.022)
Education, secondary or tertiary (d)* hurricane (d)						0.041
						(0.047)
Household shocks						
Death/illness 2003-2005 (d)						−0.032**
						(0.016)
Education, secondary or tertiary (d)* death/illness 2003-2005 (d)						0.016
						(0.033)
Adverse economic shock 2003-2005 (d)						0.004
						(0.026)
Education, secondary or tertiary (d)* economic shock 2003-2005 (d)						0.006
						(0.035)
Controls	YES	YES	YES	YES	YES	YES
Estimated probability Prob(migrant 2005= 1|intention 2005= 1)	0.103	0.106	0.107	0.107	0.105	0.080
Observations	13,036	13,036	13,036	13,036	13,036	12,842

Marginal effects; Standard errors in parentheses ; (d) for discrete change of dummy variable from 0 to 1
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Controls included are age, gender, regional and geographic dummies

Figure 2: Intention and migration

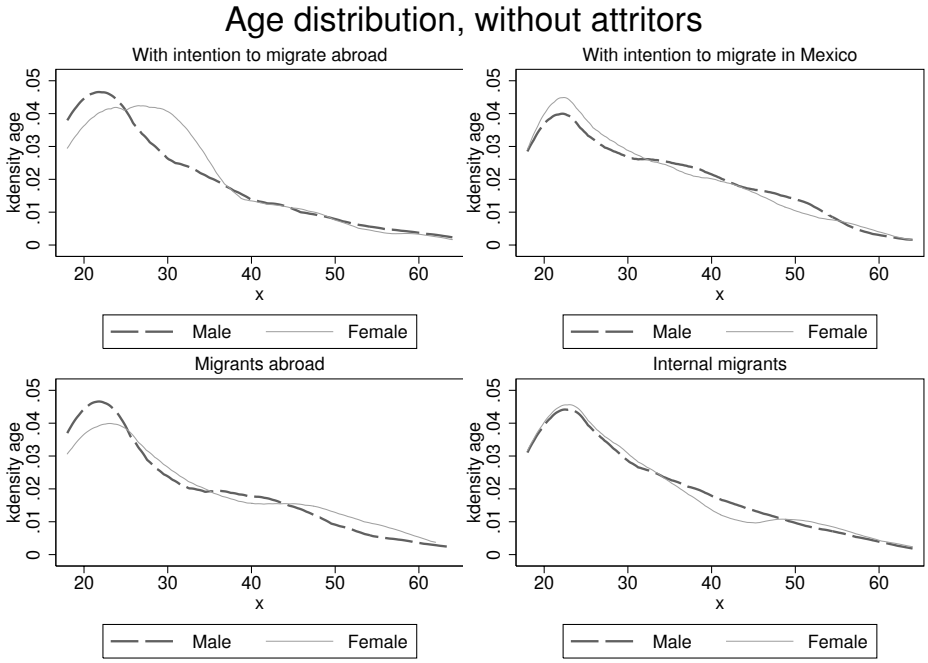
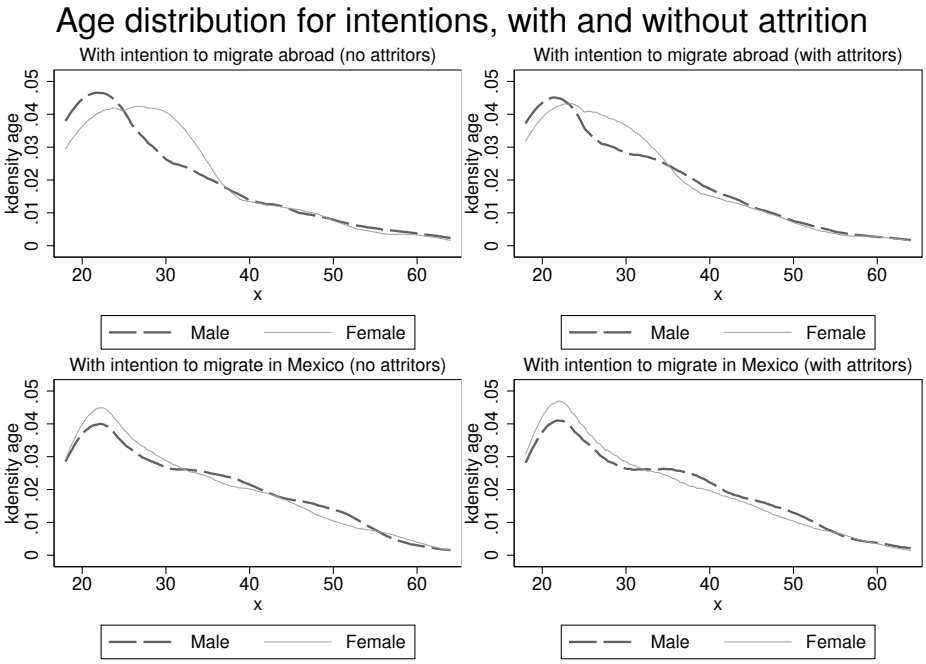


Figure 3: Robustness check for attrition



References

Ajzen, Icek. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50 (2):179–211.

Ajzen, Icek and Martin Fishbein. 1980. *Understanding attitudes and predicting social behavior*. Prentice Hall.

Balan, Jorge, editor. 1981. *Why People Move: Comparative Perspectives on the Dynamics of Internal Migration*. Paris: UNESCO Press.

Batalova, Jeanne. 2008. "Mexican Immigrants in the United States." URL `http://www.migrationinformation.org/USfocus/display.cfm?id=679`.

Bertrand, Marianne and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *The American Economic Review* 91 (2):67–72.

Binational Study on Migration. 1997. "Migration between Mexico and the United States." Tech. rep.

Borger, Scott C. 2011. "Self-Selection and Liquidity Constraints in Different Migration Cost Regimes." .

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

Borjas, George J. and Stephen G. Bronars. 1991. "Immigration and the Family." *Journal of Labor Economics* 9 (2):123–148.

Burda, Michael C., Wolfgang Hardle, Marlene Muller, and Axel Werwatz. 1998. "Semiparametric Analysis of German East-West Migration Intentions: Facts and Theory." *Journal of Applied Econometrics* 13 (5):525–541.

Cadwallader, Martin. 1992. *Migration and Residential Mobility: Macro and Micro Approaches*. Madison: University of Wisconsin Press.

Caponi, Vincenzo. 2006. "Heterogeneous Human Capital and Migration: Who Migrates from Mexico to the US?" Discussion Paper Series 2446, IZA.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. "Migration with Endogenous Moving Costs." *The American Economic Review* 86 (4):909–930.

Chiquiar, Daniel and Gordon 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).

Chiswick, Barry R. and Paul W. Miller. 2009. "The international transferability of immigrants' human capital." *Economics of Education Review* 28 (2):162 – 169.

Cornelius, Wayne A. 2001. "Death at the Border: Efficacy and Unintended Consequences of US Immigration Control Policy." *Population and Development Review* 27 (4):661–685.

Donato, Katharine M. 1993. "Current Trends and Patterns of Female Migration: Evidence from Mexico." *International Migration Review* 27 (4):pp. 748–771.

Durand, Jorge, Douglas S. Massey, and Rene M. Zenteno. 2001. "Mexican Immigration to the United States: Continuities and Changes." *Latin American Research Review* 36 (1):107–127.

Fernandez-Huertas Moraga, Jesus. 2008. "New evidence on emigrant selection." Tech. rep.

Freeman, Gary P. and Franck D. Bean. 1997. "Mexico and U.S. Worldwide Immigration Policy." In *At the Crossroads: Mexico and U.S. Immigration Policy*, edited by Franck D. Bean, Rodolfo de la Garza, Bryan Roberts, and Sidney Weintraub, chap. 1. Rowman and Littlefield, 21–46.

Hanson, Gordon H. 2006. "Illegal Migration from Mexico to the United States." *Journal of Economic Literature* 44 (4):869–924.

Heckman, James. 1974. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42 (4):679–694.

Heckman, James J. and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *The Journal of Political Economy* 93 (6):1077–1125.

Ibarraran, Pablo and Darren Lubotsky. 2005. "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census." NBER Working Papers 11456, National Bureau of Economic Research, Inc.

Kanaiaupuni, Shawn Malia. 2000. "Reframing the Migration Question: An Analysis of Men, Women, and Gender in Mexico." *Social Forces* 78 (4):pp. 1311–1347.

Liebig, Thomas and Alfonso Sousa-Poza. 2004. "Migration, Self-Selection and Income Inequality: An International Analysis." *Kyklos* 57 (1):125–146.

Lu, Max. 2005. "Do People Move When They Say They Will? Inconsistencies in Individual Migration Behavior." *Population and Environment* 20 (5):467–488.

Magnac, Th. 1991. "Segmented or Competitive Labor Markets." *Econometrica* 59 (1):165–187.

McKenzie, David, John Gibson, and Steven Stillman. 2009a. "How Important is Selection? Experimental Vs Non-experimental Measures of the Income Gains from Migration." *Journal of the European Economic Association* forthcoming.

———. 2009b. "Migration and Mental Health: Evidence from a Natural Experiment." *Journal of Health Economics* 28 (3):677–687.

McKenzie, David and Hillel Rapoport. 2007. "Self-selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks." *World Bank Policy Research Working Paper 4118* .

McKenzie, David John and Marcin J. Sasin. 2007. "Migration, Remittances, Poverty, and Human Capital: Conceptual and Empirical Challenges." Tech. Rep. 4272.

Moore, E.G. 1986. "Mobility Intention and Subsequent Relocation." *Urban Geography* 7:497–514.

Orrenius, Pia M. and Madeline Zavodny. 2005. "Self-selection among undocumented immigrants from Mexico." *Journal of Development Economics* 78:215–240.

Pugatch, Todd and Dean Yang. 2011. "The Impact of Mexican Immigration on U.S. Labor Markets: Evidence from Migrant Flows Driven by Rainfall Shocks." Tech. rep.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2):135–146.

Rubalcava, Luis and Graciela Teruel. 2006. "User's guide for the Mexican Family Life Survey first wave." Tech. rep.

———. 2008. "User's guide for the Mexican Family Life Survey first wave." Tech. rep.

Sjaastad, Larry A. 1962. "The Costs and Returns of Human Migration." *Journal of Political Economy* 70 (4):80–93.

Sutton, Stephen. 1998. "Predicting and Explaining Intentions and Behavior: How Well Are We Doing?" *Journal of Applied Social Psychology* 28 (15):1317–1338.

Todaro, Michael P. 1969. "A Model of Labor Migration and Urban Unemployment in Less Developed Countries." *American Economic Review* 59:138–148.

Van Dalen, H.P. and K. Henkens. 2008. "Emigration Intentions: Mere Words or True Plans? Explaining International Migration Intentions and Behavior." Tech. rep.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.