



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

**Migration Costs and Networks:
household optimal investment
in migration**

Andrea Gentili

Quaderni - Working Paper DSE N°867



Migration Costs and Networks: household optimal investment in migration

Andrea Gentili¹

University of Bologna

Abstract

International migration is an expensive form of investment, that only households relatively better off can afford. However poorer households have the higher incentive to migrate. Migration decision is conditional on the entry cost, expected returns and risks of migration. This paper, using data from Mexican rural and urban areas, examines the relation between household and community networks and costs and risks of migration focusing on the optimal investment in migration. To investigate an household optimal number of migrants this paper introduces a Three Step procedure to solve simultaneously for the endogeneity of network size and possible selection of migrants. The analysis confirms the inverted U-shaped relation between wealth and migration, stressing the importance of networks particularly in facilitating the migration of social strata belonging to the left tail of the income distribution. Moreover, in presence of sunk costs and/or high initial investment, household and community networks accomplish different functions.

JEL classification: O15; J11; J61

Keywords: Migration; Network effect; Costs

Introduction

This paper aims to fill a gap in the literature on the determinants of migration: the optimal number of migrants in household migration strategy. While at a first glance it could appear of marginal importance, this is a key issue in many local and international migration flows.

¹ Mail: andrea.gentili12@unibo.it ;

Permanent Address: Viale Abramo Lincoln n°6, 40139, Bologna, Italy. Tel.: +39 347 6075831

Paper presented at UNO Cumbre Ollas 2010, Omaha (Nebraska, US) and NORFACE UCL 2011, London (England)

To estimate the optimal number of migrants in household strategic behavior requires to reconsider costs and risks of migration, and therefore their relation to networks. Cost and risk of migration are not merely a barrier to entrance, but they differentiate among household and communities determining both the probability of, and the optimal investment in migration. Starting from the idea that wage gaps² are the fundamental elements driving labor migration in a household income maximization decision, this study analyses the link between budget constraints, costs and risks of migration and migrant networks.

Households are decision makers endowed with production capabilities (labor, land and financial assets) that can be allocated to improve household wealth in a world where international mobility of labor is feasible. To identify households as decision makers optimizing their wellbeing implies to determine both the probability of migration, and the optimal number of migrants (i.e. the optimal investment). Despite the little attention devoted by economic literature to the optimal number of migrants, the dimension of the phenomenon is central to improve our understanding of migration flows. For instance in our sample, the 24.70 percent of the 2024 Mexican households involved in migration has more than one member abroad. If we were computing only migration propensities and use them to forecast Mexican migration to the US, we would underestimate the flow by 756 units (37.0 per cent).

Since the seminal work of Roy (1951), many economists have investigated the causes of labor mobility. Several prominent scholars have concentrated their analysis on the relation between migration expected returns, costs, risks, networks and social capital. Borjas (1994, 1995a, 1995b), reviewing the literature on immigration to the US, focused on the quality of migrants, their wage convergence path, their contribution to the welfare state, and second-generation migrants. Ghatak et al. (1996, pag:1), presenting “a critical survey of theories of migration, their welfare and policy implication and their empirical relevance”, show that international labor migration is not the immediate response to wage differentials. Massey et al. (1993) provide the most complete review of migration theories, carefully labeling them in eight different groups, discussing their empirical evidence, pros and cons, and promoting a process of convergence. More recently, Hatton and Williamson (2003), summarizing the literature, look specifically at empirical studies on the economic and demographic fundamentals driving world migration, whereas de Haan (2006) analyses the literature on migration and its links with development studies. Lastly, Radu (2008) reviews migration literature, looking specifically at the effect of social interactions on migration and how they have been treated in theoretical and empirical research.

The relation between wealth and migration is the starting point for economists in studying the determinants of migration flows. For neoclassical theory³, migration is the result of the aggregation of rational choices made by single potential migrants trying to maximize their income in response to wage gaps across countries. The rationality of their choices and the possibility of not undertaking migration lie at the basis of voluntary labor migration (Sjaastad, 1962). However, if this were the whole truth, we should observe much larger migration flows than those observed in reality. For example, Clemens et al. (2008) showed that the yearly net return to migration from Mexico to the U.S. in 1994 was of the order of 15,000 US\$, while the cost of a *coyote*⁴ was 619 US\$ (Orrenius, 1999).⁵ Hanson (2006) estimates that, in 2000, a

² It is not one of the purposes of this paper to examine the causes behind these wage gaps.

³ Roy (1951), Sjaastad (1962), Todaro (1969), Harris and Todaro (1970), Borjas (1985, 1987, 1990, 1991), Chiswick (1999), Chiswick and Miller (2002, 2006)

⁴ A smuggler; Hanson and Spilimbergo (1999) find a strong negative correlation between attempted illegal migration and Mexican wages.

23-27-year-old Mexican migrant with 5-8 years of schooling would have covered the cost of crossing the border in 313 hours of work in the U.S.

A partial explanation of these discrepancies between theoretical predictions and empirical observations is due to the presence of borrowing constraints. Costs, thus budget constraints are in fact one of the main elements limiting migration. Nonetheless, budget constraints and imperfect financial markets can only explain why we observe relatively low levels of migration even in the presence of large wage gaps (Hatton and Williamson, 1992), but they cannot explain other empirical findings, such as the high ethnic concentration of migrants in some specific areas (friends and relatives effect).

New Economy of Labor Migration (NELM),⁶ finding in the imperfection of insurance and credit markets the main causes of migration, provides a partial explanation to these problems. Identifying the household as the decisional unit, NELM allows potential migrants to exploit a larger set of optimization strategies; in particular, migration is the result of a process of income risk minimization. Households, composed of a certain number of members, permit strategic allocation of workers in different sectors of the economy or in different countries. If risk minimization is the only objective function, we should observe widespread migration, with migrants from the same household working in different countries or economic sectors, whereas migrants usually tend to concentrate in specific groups and economic sectors.

Although NELM explicitly identifies in household internal links one of the key aspects in migration decisions, households do act independently of each other, and equilibrating mechanisms are determined by aggregate behaviors.⁷ This is not likely to be the case in the real world, where interactions outside the households have been shown to be crucial in many economic decisions,⁸ specifically on the decision to migrate.

Network theory (Fawcett, 1989; Massey and Espinosa, 1997) based on social interactions explains the high ethnic concentration of migrants and the presence of migrant flows with preferential receiving counties. Each migrant, creating new links in the receiving country and retaining⁹ some in the sending country, modifies the social environment in both, allowing the accumulation of migration-specific knowledge (migration social capital) able to reduce the costs and risks of migration and generating a self-perpetuating mechanism. In particular, networks affect the relation between migration and wealth, mitigating the effect of budget constraints not only by reducing costs and risks, but also acting as substitutes for financial and insurance markets (Yang, 2008).

The endogenous process identified by network theory is not limited to potential migrants, since migration alters the whole sending country's socio-economic environment. The accumulation of migration-

⁵ Using MMP data, Orrenius (2001) showed that, during the period 1978-1996, around 69 percent of Mexican migrants to the U.S. passed the border by hiring a coyote for average prices of US\$385 – 715; Cornelius (2005) found that, after the change in U.S. immigration policy after 9/11, the cost of hiring a coyote increased by around 37%

⁶ Stark and Levhari (1982), Stark (1984, 1991) Taylor (1986)

⁷ Each migrant reducing the supply of labor in the sending country and increasing it in the receiving country, increases wages in the sending country and decreases them in the receiving country, respectively. Similar aggregate behaviors happen in markets different from that of labor.

⁸ See for example Goyal (2007).

⁹ Hanson and Spilimbergo (1999) find that the purchasing power of both the U.S. and Mexico matters in border apprehension, suggesting that potential illegal migrants expect to retain connections in both countries.

specific social capital, remittances, changes in the distribution of wealth¹⁰ and land, all concur in generating a new set-up which has the potential to produce favorable new conditions for migration. Cumulative causation, as developed by Massey and followers, goes in this direction, providing a general framework for tracing potential migration paths.

Household and community networks channel international mobility of labor, however they are determinant only if we consider them in conjunction with cost and risks of migration. This complex endogenous relation is one of the most challenging empirical issues associated with migration studies. The other main problem is the sample selection. At priori we cannot exclude that there are some unobserved characteristics driving the propensity of certain household to migrate.

Since the aim of this paper is to analyze both the propensity and the optimal number of migrants in a single framework, it is necessary to develop an empirical approach able to disentangle simultaneously sample selection and endogeneity of network size. Usually, to tackle selection, the Heckman correction method is applied. Since the HTS procedure is not reliable in the presence of endogenous phenomena, following Mroz (1987) and Piacentini (2008), this paper introduces a three-steps procedure new for international migration analysis. This approach allows to identify the effect of selection (Probit and IV-Probit) and to examine the number of migrants a household sends abroad (IMR and IV-TSLS).

Using the Mexican Migration project data (MMP), the econometric results presented in this paper show that migration and wealth are non-linearly related (inverted U-shape) and that household and community level networks are complements in the migration decision, but substitutes in the optimal number of migrants. As pointed out networks have a role in migration decision conditional on the presence of costs and risks of migration. The differentiated structure of costs and risks implies that different networks accomplish different functions. This result, new in empirical literature, requires to re-evaluate the relation between networks and migration costs and risks.

Section 1 Basic Model

Consider a dual economy along the lines of McKenzie and Rapoport (2007) with $i = 1, 2, \dots, I$ households of size $l_i = 1, 2, \dots, L$.¹¹ Households are endowed with an illiquid asset A_i which is the basis of the family business. Let us assume, for the sake of simplicity, that income is equally shared among household members, that each household member lives for two periods $t = 1, 2$, and that each member provides one unit of work to the family business. The marginal product of family business is also linearly increasing in the initial endowment and marginally decreasing in the number of workers, so that one simple possible representation is:

$$y_i = A_i l_i - \frac{b l_i^2}{2} \quad (1)$$

¹⁰ Docquier and Rapoport (2003) develop a model to analyze the link between migration, remittances and inequality. The main pro of their model is to take into account the effect of migration in local labor markets, making migration an endogenous process even in the absence of networks able to reduce the cost of migration.

¹¹ Household size is treated as a continuum number.

A household member can migrate at a cost c_i , which is assumed to be fixed and exogenous. If a household member migrates he or she will receive a salary $w_f \sim U[\mu_{w_f}, \sigma_{w_f}]$, where f stands for foreign country. We assume that risk¹² is fixed and exogenous, and that μ_{w_f} is equal for each household member and for each potential migrant across the society.¹³

In order to take into account the incompleteness of insurance and financial markets we assume that borrowing is not allowed,¹⁴ so that migration must be financed through savings. This implies that migration is impossible in the first period, and that the decision to migrate is the effect of household income maximization in the second period, when the household can use savings to finance migration. The household cannot save all its first period income, but it needs to consume at least S (Subsistence need) for each member at that time.

We also assume that $\mu_{w_f} > S$ and that $A_i - \frac{bl_i}{2} > S$. The first assumption implies that migration is appealing, and the second that the share of wealth each household member has is large enough to ensure survival.

In this simplified framework, a household¹⁵ chooses the share of members who migrate m_i , to maximize the second-period income. Because income maximization does not take into account risk and uncertainty aversion,¹⁶ in order to implement them this model assumes Constant Absolute Risk Aversion (CARA) utility functions. ρ is the Arrow-Pratt measure of absolute risk aversion and expected salaries are distributed as a normal.¹⁷ Making use of CARA properties, and assuming no discounting, the household second-period maximization is:

$$\max_{0 \leq m_i \leq 1} A_i l_i (1 - m_i) - \frac{bl_i^2 (1 - m_i)^2}{2} + m_i l_i \mu_{w_f} - m_i l_i c_i - \frac{\rho}{2} m_i^2 l_i^2 \sigma^2 \quad (2a)$$

S. t.:

$$A_i - \frac{bl_i}{2} - m_i c_i \geq S \quad (2b)$$

To simplify notation, we suppress subscript i . Solving first-order conditions, m^* , the optimal rate of outmigration is:

$$m^* = \frac{-A + bl + \mu - c}{l(b + \rho\sigma^2)} - \frac{\lambda c}{l^2(b + \rho\sigma^2)} \quad (3)$$

¹² Risk is here intended as the risk of failure in migration and as uncertainty about future income, due to lack of information of the receiving country's labor market.

¹³ This to simplify the analysis; in principle it may be extended by introducing various expected incomes and costs based on household social strata. This would imply the introduction of different networks which might have different properties, complicating the theoretical analysis and making an empirical investigation difficult or even unfeasible.

¹⁴ Thus the focus is on migration between countries with different levels of development (i.e., South-North migration).

¹⁵ How the decision is taken inside the household is not examined here.

¹⁶ This is the first difference with respect to McKenzie and Rapoport (2007), who do not take into account risk/uncertainty.

¹⁷ Salaries are usually distributed as a log-normal distribution, however this assumption is technically necessary, and it does not determines our results being the investigation of risk beyond the scope of this paper.

m^* is increasing in expected salary μ , and decreasing in its variance $\sigma^2 \forall \sigma > 0$ and in migration costs c . λ is equal to 0, unless the vinculum binds. If it binds, as in McKenzie and Rapoport (2007), the constrained migration rate is:

$$\tilde{m} = \frac{1}{c} \left(A - \frac{bl}{2} - S \right) \quad (4)$$

We can compute the highest level of fixed assets at which a household has no possibility of migrating, $\underline{A} = \frac{bl}{2} + S$. Define A_1 the level of illiquid assets, above which a household is no longer trapped by subsistence needs:

$$m^* = \tilde{m} \Rightarrow A_1 = \frac{2\rho\sigma^2 l(S+l) + 2c\mu + 2c(2bl-c) + bl(2S+bl)}{(2l\rho\sigma^2 + 2bl + 2c)} \quad (5)$$

If either $\sigma = 0$ or household is risk neutral, the analysis comes back to McKenzie and Rapoport (2007). We can also identify the minimal value of A at which households will not choose migration as the optimal behavioral strategy. This value is equal to that computed by McKenzie and Rapoport (2007):

$$\bar{A} = bl + (\mu - c) \quad (6)$$

The migration rate changes depending on the initial endowment:

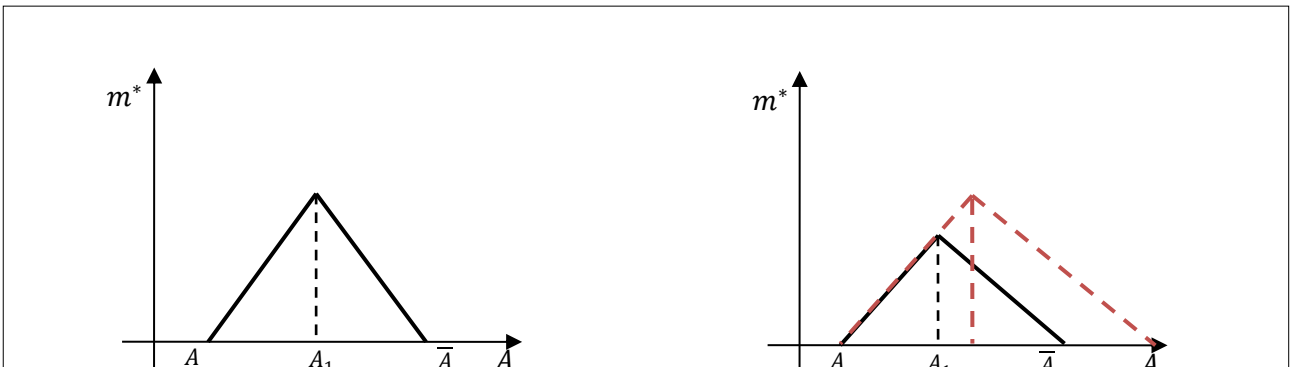
$$\frac{\partial m^*}{\partial A} = \begin{cases} -\frac{1}{l(\rho\sigma^2 + b)}, & \lambda = 0 \\ \frac{1}{c}, & \lambda > 0 \end{cases} \quad (7)$$

Thus, the migration rate path in relation to wealth levels is:

$$m^* = \begin{cases} 0 & , & A \leq \underline{A} \\ \frac{1}{c} \left(A - \frac{bl}{2} - S \right) & , & \underline{A} < A \leq A_1 \\ \frac{bl - A + \mu - c}{l(b + \rho\sigma^2)} & , & A_1 < A \leq \bar{A} \\ 0 & , & \bar{A} < A \end{cases} \quad (8)$$

This system can be graphically represented as in McKenzie and Rapoport (2007) though a “triangular” representation of the relation between the share of migration and the initial endowment of immobile capital. Introducing the risk/uncertainty of migration generalizes the model, so that McKenzie and Rapoport (2007) is a specific case of this model.

As shown in Picture 1, the household migration rate is a triangular function of immobile assets, migration is 0 under subsistence needs when the initial endowment of assets is below \underline{A} . First it increases with wealth up to A_1 , and then it decreases until it returns to 0 when the initial endowment is above \bar{A} .



Section 2 Data

In order to investigate the optimal investment in migration, thus the optimal number of members sent in migration, the ideal dataset should include individual and community information on household resources endowment (labor, land and financial assets), household histories of migration, and community migration capital. I focused on the MMP118 database, a collaborative research project based at Princeton University and the University of Guadalajara¹⁸organizing information on 118 communities surveyed in the period 1987-2007. In my opinion and to the best of my knowledge, this is the database closest to the ideal one.

There are several important pros in using Mexican data. First of all Mexico and US share the longest border between a developing and a developed country. Secondly Mexico represents a promising setup to test the quality of the empirical procedure proposed because migrants networks are diffused and stratified spatially and in time. Additionally the availability of a large number of studies focusing on the determinants of Mexican migration to the US allows to compare result with a well-established background.

The dataset includes information on 19,726 Mexican households. It is a household-level database containing information on household composition, economic and migratory activities of household members, land ownership and usage, home/real estate ownership and amenities, vehicle and livestock ownership and financing, and business ownership and operations.

¹⁸ More information about the MMP database can be found in MMP website: <http://mmp.opr.princeton.edu/>

Although the MMP probably represents the most comprehensive database on Mexican migration, it includes no reliable information on household income, so that an alternative measure had to be found. The MMP includes information on household access to infrastructures, such as access to electricity and running water, dirt or tile floors, and household ownership of some durable assets such as cars, radios and television sets, allowing the application of PCA to derive a household wealth indicator. Filmer and Pritchett (2001) used PCA to derive household wealth in India and several other countries, and showed that an asset-derived index is as accurate as information on expenditures in predicting school enrollment of children. McKenzie (2005) showed how information from the MMP can be used in conjunction with national income and expenditure surveys (ENIGH) to predict non-durable consumption (NDC) and derive a reliable inequality index for Mexico. Since the investigation of inequality is beyond the scope of this paper, I focus on Principal Components.¹⁹

2.1 Household Composition

In the present analysis a household is defined as migrant when one or more of its members have migrated to the U.S. in the three years prior to the survey. According to this definition, in the sample there are 2,024 migrant households (15.81 percent). Household heads with previous migration experience are not dropped from the sample,²⁰ expecting previous migration experiences to play a key role in developing a network.

Human Capital Assets identify a series of structural household elements which are likely to affect migration, as shown among others by Winters et al. (2001).

The size of migrant households (that is, the number of household members) is larger than the size of non-migrant households by around 0.8, and the difference is statistically significant (0.01 confidence level) when a t-test is performed.²¹ The average age of migrant household heads is below that of non-migrant household heads.²² As well established in literature, it is more likely that subjects migrate for the first time when they are relatively young, to maximize expected returns.

In line with previous findings, the *education level* of non-migrant households is higher than that of migrants.

¹⁹ Using assets as instruments for household wealth overcomes two main problems typical of income or expenditure data, since these are more subject to measurement errors. Secondly, consumption expenditures and income need to be normalized, to take into account the number of members in the family, whereas the utility of assets is usually the same for all household members, independently of their number.

²⁰ McKenzie and Rapoport (2007) only study first time migrant households.

²¹ Winter et al. (2001) report a difference of 3 members in favor of migrant households, but their measure only refers to the number of adults.

²² Only Winters et al. (2001) find migrant household heads to be older than non-migrants.

Variable	Table 1: Descriptive Statistics						
	Migrants		Non-Migrants		Total		T-Test
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Sample Size	2024	-	10781	-	12805	-	
Number of recent US Trips	1,374	0,787	0,000	0,000	0,217	0,591	*
Human Capital Assets							
N° of Members	5,399	2,527	4,629	2,285	4,751	2,342	*
N° of Workers	2,216	1,516	1,758	1,200	1,830	1,266	*
Percentage of Males	0,493	0,190	0,452	0,212	0,458	0,209	*
Household age	43,482	13,881	48,388	15,587	47,612	15,434	*
Education Level	5,073	3,438	5,669	4,230	5,575	4,121	*
Cross effect Wealth and Education							
educ*wealth	26,350	19,599	30,013	25,016	29,434	24,276	*
Physical Assets PCA							
Wealth	5,028	0,848	4,968	1,039	4,978	1,011	*
Household Network							
Historic Migration Experience	4,200	6,504	0,723	2,259	1,273	3,548	*
Current Network	16,417	21,445	7,702	14,028	9,080	15,762	*
U.S. resident	1,063	1,646	0,463	1,092	0,557	1,217	*
Community Network							
Migration Prevalence	0,268	0,139	0,189	0,148	0,201	0,149	*
Migration Prevalence*Wealth	1,346	0,752	0,928	0,747	0,994	0,763	*
Physical Costs							
Average Distance from the U.S.	1901,343	198,474	1889,724	279,267	1891,560	268,148	
Border (dummy)	0,041	0,197	0,125	0,331	0,112	0,315	
Economic Indicators							
Mexican Unemployment Level	0,034	0,008	0,034	0,008	0,034	0,008	
US Unemployment Level	0,057	0,008	0,055	0,008	0,056	0,008	
Exchange Rate	5,339	3,092	6,573	3,233	6,378	3,242	

Table 1: Descriptive statistics. Source: authors from MMP118, NATLHIST, NATLYEAR, Google Maps Tools for distances

2.2 Household Capital Assets

Since income information is not reliable, it is necessary to identify an alternative way of measuring household wealth. Following Filmer and Pritchett (2001) and McKenzie (2005), a reliable index of household wealth is obtained by using PCA on a set of information on household facilities and asset indicators such as land holdings, house building materials and amenities. The basic idea behind PCA is to

describe a multivariate dataset in the most “simple” way possible through a set of derived uncorrelated variables, each of which is a linear combination of the variables in the original dataset.

PCA makes one stringent assumption: *linearity*, identifying the combination of original basis which best represent the dataset. The First Principal Component is the linear combination of all the variables which capture the largest variability and thus the largest amount of information.²³

Assuming that what mainly determines variations in housing construction materials, amenities, vehicle ownership and business holdings is wealth,²⁴ the first factor (Principal Component) identifies the wealth level of a household. MMP118 includes 27 asset indicators grouped in four main categories: farming and breeding, property holding, household amenities, and business holdings. Table 2 lists the scoring factors of each group and all components.²⁵

Factors derived using information about housing and amenity ownership are highly correlated with the *Total Index 25 (TI25)*. *TI25* is similar to the wealth index derived by McKenzie (2005) and used in McKenzie and Rapoport (2007). Only three elements differ. *TI25* includes the number of hectares owned and the number of businesses held by the household not taken into consideration in McKenzie and Rapoport (2007), whereas information about education is excluded from the PCA index. The introduction of the first two elements aims at increasing the number of variables which may, in principle, reflect the long-run wealth level of the household. Although educational attainment is highly correlated with the wellbeing of a household, the information is not used to derive *TI25*, since it is a regressor for migration analysis.

The farming factor has a low and negative correlation with *TI25*. One explanation is that, on average, rural households are poorer. The very low coefficient may be explained by the large amount of 0's in farming and breeding activities: only around 2000 households own land and even fewer own animals. Similarly, the low correlation coefficient between the *Business* factor and *TI25* may be explained by the large amount of 0's in the sample.

²³ See Filmer and Pritchett (2001) for exhaustive information.

²⁴ It is implicitly assumed that everyone prefers higher (in quality or numbers) asset ownership than lower ones. Differences are not the effect of tastes but of different economic opportunities.

²⁵ For what concern housing, business and farming activities the information refer to prior the last migration was undertaken if the household is defined as migrant. This was not possible for amenities.

Table 2: PCA; Scoring Factors					
	Farming PCA	Housing PCA	Amenities PCA	Business PCA	Total Index 25
<i>Farming and Breeding</i>					
Land Ownership	0.8533				-0.0998
N° of hectares per household	0.2968				0.0138
Machinery	0.7934				
Fertilizers	0.8627				
Insecticides	0.8399				
Cows	0.322				
Pigs	0.1129				
Horses	0.3927				
Burros	0.3312				
Oxen	0.1827				
Chicken	0.3025				
<i>Housing</i>					
N° of Property Holding		0.1825			0.2088
Construction1: adobe and tile roof		-0.4428			-0.437
Construction2: brick and tile roof		-0.5786			-0.3448
Construction3: brick and cement roof		0.7952			0.6005
Construction4: wood		-0.1013			-0.0902
Floor1: dirt		-0.4046			-0.5386
Floor2: cement		-0.6122			-0.3832
Floor3: finished		0.8302			0.6752
N° of rooms		0.5546			
N° of rooms/member					0.3445
<i>Amenities and vehicles</i>					
Running water			0.4162		0.3807
Electricity			0.4084		0.3493
Sewer			0.5326		0.5473
Stove			0.5699		0.5224
Refrigerator			0.7419		0.7142
Washing machine			0.6849		0.6397
Sewing machine			0.461		0.4194
Radio			0.39		0.3197
TV			0.5749		0.4948
Stereo			0.6241		0.5879
Phone			0.5877		0.5938
Car			0.422		0.4116
Van			0.3239		0.2925
Bus			0.069		
Tractor			0.0744		
Taxi			0.0449		
Motorcycle			0.0493		
Other vehicle			0.0055		
<i>Business Holdings</i>					
N° of business holdings				0.9956	0.1931
Business type: store				0.4652	
Business type: street vendor				0.395	
Business type: restaurant/bar				0.2258	
Business type: workshop				0.3541	
Business type: factory				0.1068	
Business type: middleman				0.1971	
Business type: personal service				0.1636	
Business type: professional service				0.1112	
Business type: other service				0.0971	
Business type: agriculture				0.2814	
Business type: cattle raising				0.242	
Business type: other				0.3769	
Eigenvalues associated with first component	3.40048	2.742	3.69581	1.91871	5.01601
Share of variance associated with first component	0.3091	0.3047	0.2053	0.1476	0.2006
Number of variables used	11	9	18	13	25
Correlation with T125	-0.0492	0.7881	0.935	0.1931	

Table 2. Author from database

2.3 Network Variables

Household network is identified by three main variables: the number of *historic migration experiences*, *current network*, and household relationships with *U.S. residents*.

The *historic migration experience* is defined as the sum of the number of migration experiences of household heads and/or their spouses, sons and daughters prior to the last 3 years.

Current network is the number of friends and relatives, not belonging to the household but to the extended family, who were abroad in the year of the survey.

U.S. resident is the number of relatives actually residing in the U.S.. Both variables are likely to affect migration decisions. Each household member can, in fact, receive financial support, assistance and information from the network. Specifically, three aspects of current household migration networks may have a great influence on migration decisions: financial support, housing, and information.

Financial support may be fundamental in overcoming budget constraints, particularly when crossing the border is expensive. Those who have already migrated can finance migration of co-villagers and relatives for various reasons: altruism, inequality aversion, social norms, loan repayments, household income maximization strategies, household income risk minimization, or speculation. Independently of the reason, financial support allows potential migrants to overcome budget constraints.

Housing has been shown to represent the main cost, at least in the first phase of residence in the receiving country. The larger the number of connections (or, even better, some residents), the greater the reduction of this cost. Housing support is usually “rent-free”.

Last but not least, information: there are two main sources of risk, border crossing and unemployment. To be in contact with someone who has recently migrated or who is currently residing in the U.S. can greatly improve the information available to potential migrants and their households. Specifically, recent migrants can provide information on how to cross borders, can introduce potential migrants to *coyote*, or help newcomers with bureaucracy. Contacts in the receiving country can provide information on potentially available jobs.²⁶

Community-level networks are sources of information and a series of services which could reduce both the psychological and the physical cost and risks of migration, for example providing ethnic goods through formal or informal markets, organizing money transfers and transportation services, spreading knowledge on the receiving country, and even running development programs in sending communities. However, identification of community-level network effects on migration is not straightforward.

The main difficulty is finding an appropriate measure able to capture all these functions. Following Durand et al. (2001), I used *migration prevalence* ratios to incorporate the community-level migration network: *migration prevalence* does not describe the migration flow per se, but rather, a phase in its development. It is a useful measure which can capture the level of development of migration flows in a

²⁶ In Texas and California resident migrants often act as recruiters in seasonal farming work. This may explain the high ethnic concentration of Mexicans in certain economic sectors. Moreover, some migrants resident in the U.S. become entrepreneurs, hiring compatriots as employees.

community. *Migration prevalence* ratios are calculated with information on the date of birth of household members and the year of their first trip to the U.S..²⁷

Nevertheless, *migration prevalence* has some important disadvantages. First, as already highlighted by Durand et al. (2001), “it tends to dehistoricize migration”. This means that specific local or global events affecting migration rates (e.g., the 1925 railway construction, the 1940s Bracero Program, the 1980s economic crisis, 9/11) may occur at any moment in the history of migration of different communities, thus implying different effects. The use of IV, as in McKenzie and Rapoport (2007), not only solves endogeneity but also avoids the *dehistoricization* of migration.

Migration prevalence may be biased upwards or downwards, depending on internal and permanent migration dynamics. It may be overestimated if internal and international migrations are substitutes, or if migration becomes permanent (migrant households disappear from surveys).

Migration prevalence is a very powerful tool when applied to migration contexts presenting a prevalent destination country, such as the U.S. for Mexico. At the same time, *migration prevalence* captures some innate propensities of certain communities to migrate. This propensity cannot be captured with variables such as the stock of compatriots in the receiving country, so that migration prevalence improves understanding of the migration flow, allowing better generalization network effects derived from individual data.

2.4 Physical Costs and Economic Indicators

*Distance*²⁸ is likely to be an approximate measure of the cost of migration. In cross-country migration flow analysis, distance is always used as a rough measure of physical cost of migration and of cultural distance. In unidirectional analysis, distance should still have a negative significant influence on the decision to migrate, if the sending country is large enough and if the cost of moving between countries is high enough to become a barrier.

Economic Indicators are likely to affect household migration strategies. Average levels are computed as the means in the last three years before the survey was undertaken. *Unemployment* levels are likely to be a partial measure of risk. To avoid possible fluctuations due to the business cycle I used the difference in unemployment levels between U.S. and Mexico and the *exchange rate* to measure the monetary advantages of migration. It affects not only the expected return of migration, but the family income risk minimization strategy. Migrating when exchange rates are high increases household wellbeing in Mexico. Remittances have higher value, since they are in U.S. dollars. Besides, having remittances in U.S. dollars ensures households against hyperinflation and monetary devaluation.

2.5 U.S. Immigration Policy

One of the aims of this research is to improve the methodological approach in the study of the determinants of migration, to provide policy-makers with better forecasting instruments. Of particular

²⁷ For a complete explanation on how to compute migration prevalence ratios, see Massey et al. (1994).

²⁸ Distance is measured as the average distance of the Mexico state capital of the community in question and the U.S. state capitals of California and Texas.

relevance from the policy maker point of view is the effect of legislation on potential migrant decisions. In this analysis the implementation of U.S. immigration legislation is made using the index proposed by Ortega and Peri (2009).²⁹ Authors database contains information on the immigration legislation of 15 OECD countries over the period 1980-2005. Each change in legislation is associated with a (+1) or a (-1), whenever a reform increases or decreases the tightness of immigration laws. Since Mexican emigration is mainly unidirectional to the U.S., this work focuses on U.S. legislation over the period 1987-2005 (see Table 3).

The database provides three variables on the tightness of entry, stay and refugee regulations. The focus here is not on refugees, so the refugees variable is not taken into consideration; moreover, since it is not within the scope of this paper to examine U.S. immigration policy in detail, I produced a single index, which is the yearly mean of entry and stay. Using only one index for the tightness of U.S. immigration policy also saves degrees of freedom in “temporal” analysis. The resulting variable (*avglaw*) is expected to have a significant negative influence on migration.

Table 3: US immigration Policy						
Variable	Migrants		Non Migrants		Total	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
US Immigration Policy (Peri)						
Entry	-3,2488	1,6587	-3,5589	1,2205	-3,5099	1,3045
Stay	-1,1525	0,6517	-1,1574	0,7139	-1,1567	0,7044

Table 3: Authors from Ortega and Peri (2009)

Section 3 Empirical Specifications

In order to analyze the optimal investment in migration conditional on the probability to be involved in migration, it is necessary to control simultaneously for both migrant selection and endogeneity of network size. This paper introduces a procedure based on the three-step procedure proposed by Mroz (1987): the first step solves the self-selection problem by examining the dichotomous choice of migration; the second step tackles potential endogeneity in network size by using instrumental variables; the third step identifies the network effect by including both sample selection and instrumental variable approach in a structural equation for the number of migrants.

To show the efficiency of this approach, estimations for the HTS procedure and the IV approach are provided. The HTS procedure was used, among others, by Winters et al. (2001), and the IV-Probit approach was successfully applied by McKenzie and Rapoport (2007).

²⁹ For more details see the Giovanni Peri website <http://www.econ.ucdavis.edu/faculty/gperi/> and Ortega and Peri (2009).

3.1 Solving Sample Selection, HTSp

The reduced form of the econometric model for the decision to migrate (d) can be formulated as follow:

$$d^* = W\alpha + u_1 \quad (9)$$

$d = 1$ if $d^* > 0$, 0 otherwise. The level of migration equation is:

$$l = W\beta + u_2 \quad (10),$$

observed if $d = 1$, with $(u_1, u_2) \sim N(0, 0, 1, \sigma_2, \rho)$ and $W = \{L, Z_n, Z_c, n, N, nN\}$

The estimation is made with the HTS procedure: the first stage is estimated by maximum likelihood probit, and the second stage by a truncated least squares regression. This procedure computes unbiased estimators in the second stage (by including the IMR), but it cannot solve the endogeneity problem.

As pointed by Winter et al. (2001), if there were no entry costs, the household income maximization decision would be the same as the decision of the level of migration. However since migration is costly, the decisions are different. The optimal number of migrants is in fact independent of the fixed cost, while undertaking migration a household requires positive returns, including the fixed cost of that migration. While Winter et al. (2001) had no measure for costs, so that the selection equation lacked an identification variable, two rough measures for the migration cost are available here: *distance* and *border*.

Table 3 reports estimation results. Robust standard errors were obtained by the bootstrapping method. Appendix report HTSP with non-robust, robust and cluster robust standard errors.³⁰ Although estimations cannot be directly compared with those by Winters et al (2001), the two analyses produced similar results, highlighting the importance of network and wealth variables.

The main variables of interest, those concerning wealth and networks, are all significant and have the expected direction. *Wealth* has a positive effect on migration decisions, whereas the negative sign of squared wealth suggests that migration propensity decreases after a certain threshold is reached. Thus, as hypothesized, there is an inverted U-shape relation between wealth and migration. While wealth affects the migration propensity, it has a slightly significant positive effect on the optimal number of migrants. This is in line with the initial investment problem.

As expected, *migration prevalence* has a statistically significant and positive effect on migration. Similarly, *household migration experience*, *current network*, and *U.S. resident* all positively affect migration. While *migration prevalence* and *U.S. resident* affect only the probability of migration, past migration experience and *current network* also positively affect the number of migrants. The positive and significant effect of this group of variables confirms the positive influence of migrant networks, on both the decision to migrate and the number of migrants sent abroad. Nonetheless, the positive significance of all the estimated coefficients highlights that community and household-level networks are both important in the migration decision. Thus, they are, at least partly, complements. In addition, community-level networks

³⁰ Robust standard errors were applied. Differences in SE are low (below 10-15%) between robust and non-robust estimates for all significant variables. Cluster robust SE show greater differences, particularly as regards migration prevalence. Clustering was made at community level and this seems to be imprecise: the sample probably contained heteroskedasticity at community, state and national level.

affecting only the probability of migration and not its optimal level, convey forms of information and support that are different from household ones.

In contrast with the findings of Winters et al. (2001), there is no evidence of education effect on either the propensity to migrate³¹ or the number of migrants. The weak negative link (0.05 significance level) between the cross-effect of education and wealth and the number of migrants corroborate the idea that education should be considered in the wealth indicator as a measure of household asset levels.

In line with previous findings,³² both the size of the household (measured as number of *workers*³³) and the proportion of males, positively affect the propensity to migrate and the number of migrants, while the age negatively affects migration.

Distance, as expected, has a negative influence on the migration decision, since it roughly identifies the cost of migration, but it is not significant at the 0.05 confidence level. *Border* has a negative influence on the migration decision. Communities belonging to *Baja California Norte*, *Nuevo Leon* and *Chihuahua* have historically lower levels of emigration to the U.S., migration being less necessary because of the greater number of U.S. firms across the border. Nonetheless, an F-test of the simultaneous significance of *Distance* and *Borders* strongly rejects the null hypothesis.

Contingent factors all affect both the migration decision and the number of migrants, and all present the expected sign. Lastly, the significantly positive value of the *IMR* is in line with expectations: households with higher values of the variables facilitating migration are those that would like to send more members abroad.

³¹ There is a significant, but very small, negative effect of squared education.

³² See Massey et al. (1994), Winter et al. (2001), Fussel and Massey (2004) and Konseiga (2006).

³³ The same effect is also found when the size of the household is expressed as the *number of members*.

Table 4: HTSP		Selection	F-Test	Numb.	Joint F-Test
		Rob. SE		Rob. SE	
Human Capital	N° of Workers	0,186 ***	0,000	0,245 ***	0,000
		0,013		0,030	
	Age	-0,033 ***		-0,013	
		0,007		0,007	
	Age Squared	0,000		0,000	
		0,000		0,000	
	Sexratio	0,363 ***		0,351 ***	
		0,082		0,097	
	Educ.	0,042		0,056	
		0,030		0,031	
Educ. Squared	-0,002	0,001			
	0,001	0,001			
Educ*Wealth	-0,011	-0,015 *			
	0,007	0,006			
Fiscal Capital	Wealth	1,112 ***	0,000	0,461 *	0,000
		0,180		0,211	
	Wealth Squared	-0,101 ***		-0,031	
		0,019		0,020	
Household Network	Hist. Migration	0,103 ***	0,000	0,049 ***	0,000
		0,007		0,012	
	Current Net	0,005 ***		0,005 ***	
		0,001		0,002	
	US Res.	0,081 ***		0,033 *	
0,010		0,016			
Community Network	Migration Prev.	1,986 ***	0,000	0,331	0,000
		0,521		0,816	
	Mig. Prev.*Wealth	-0,278 *		0,033	
		0,112		0,155	
Contingent Factors	Unem. Diff.	-3,853 *	0,000	-5,262 *	0,000
		1,765		2,532	
	Exchange Rate	-0,022 **		-0,034 ***	
		0,007		0,009	
	Avg. Law index	-0,045 *		-0,093 ***	
0,019		0,023			
Selection	Distance	0,000	0,000		
		0,000			
	Border	-0,393 ***			
Constant		-3,082		-1,247	
		0,462		0,772	
Inverse Mills ratio				0,510 **	
				0,172	
Censored obs	10781				
Uncensored obs	2024				

Table 4. Source: author computation from MMP118.

Robust standard errors are obtained by bootstrapping method; F-test shows that all the coefficient of the subgroup are 0 simultaneously

3.2 Ruling Out Endogeneity: IV Approach

Solving the problem of sample selection does not guarantee that estimates are consistent. In the present analysis three sets of instruments were used, listed in Table 4. Following Woodruff and Zentoro (2007) and McKenzie and Rapoport (2007), the first instrument is the average migration rate by state over the period 1956-59, at the peak of the Bracero Program (1942-1964).³⁴ The second instrument is the 1924

³⁴ Thanks to Professor McKenzie for providing me original data from González Navarro (1974).

migration rate by state. The two sets of instruments are explained in detail in McKenzie and Rapoport (2007). The third set combines the first two and adds information on visa accessibility and average U.S. wages in the three years prior to the survey.

Table 5: Instrument Sets								
Instrumental Variables	Mean	S.D.						
1924 State Mig. Rate	0.097	0.065						
1924 St. Mig. Rate*Wealth	0.494	0.353						
1956-59 St. Mig. Rate	0.036	0.037						
1956-59 St. Mig. Rate*Wealth	0.179	0.179						
VISA accessibility	0.064	0.042						
Log US wage last 3.	2.409	0.192						
Correlation	MP	MPW	M24	M24W	M50	M50W	VISA	LUSW
Migration Prevalence	1.000							
Mig. Prev.*Wealth	0.951	1.000						
1924 State Mig. Rate	0.322	0.350	1.000					
1924 St. Mig. Rate*Wealth	0.270	0.372	0.950	1.000				
1956-59 St. Mig. Rate	0.272	0.251	0.535	0.450	1.000			
1956-59 St. Mig. Rate*Wealth	0.266	0.308	0.555	0.544	0.954	1.000		
VISA accessibility	0.098	0.106	0.269	0.238	0.321	0.315	1.000	
Log US wage last 3.	-0.251	-0.221	-0.215	-0.166	-0.226	-0.219	-0.123	1.000

Table 5: IV Set

Correlations between instruments and instrumented variables are low, but not too low to flag a problem of weak instruments. As reported in Appendix under an IV Two-Stage Least Squares (2SLS) procedure, instruments appear to be exogenous, necessary and not weak. TLS is computed both for migration probability and number of migrants. Since the dependent variable is binary or discrete, robust standard errors were used. 2SLS estimations are not discussed here, since estimates are not unlike IV-Probit ones. Nonetheless, although 2SLS requires less structural hypothesis than IV-Probit, the binary nature of the dependent variable may lead to inconsistent estimates. IV-Probit estimates are reported in Table 5, estimation was undertaken with Newey's Two-Step³⁵ Estimator (Newey and West, 1987), since the maximum likelihood estimation could not be computed. As shown in Newey and West (1987), the two-step method estimates consistent values for parameters, but is less efficient in estimating SE in comparison with MLE. It is possible that, if the instruments are weak or too strong (in conjunction with large sample size), the standard errors may be inconsistent. The over-identification test and post-estimation analyses are made with STATA10 *overid* plugin.

A Wald test of exogeneity confirms the endogeneity problem and the need for an IV approach at 0.05 confidence interval when using set A, and at 0.01 confidence interval when using set B; it is refused at 0.05 confidence level with set C. The over-identification test (Amemiya-Lee-Newey test minimum chi squared statistic) fails to reject the over-identification restriction, corroborating the validity of the

³⁵ The name "Two-Step" oversimplifies the approach.

instruments used. Correlation matrix, 2SLS and IV-Probit suggest that the instruments are not weak but also not too strong, corroborating the idea that IV-Probit is the correct approach to use.

All the coefficients analyzed have the expected sign and are robust to changes in the instrument set. *Wealth* and migration probability have an inverted U-shaped relation, and all network variables positively affect migration. Only *current network* appears to be non-significantly different from 0 with set A and set B. The instrumented variable, *migration prevalence* and its cross-effect³⁶ with *wealth*, positively affects poorer households' decision to migrate. This is in line with the idea that networks affect more social strata with lower access to information and economic opportunities.

Lastly, the control variables, *unemployment difference*, *exchange rate* and the *law tightness*, are all non-significant or slightly-significant. When a test of joint significance is performed, they are significant ($\chi^2(3) = 13.64 - p=0.0034$). The *exchange rate* is the only variable which was never significant in all three estimations.³⁷

³⁶ However we are aware of the fact that estimation for cross variables could be non-informative in binary instrumental analysis.

³⁷ A partial explanation for these results can be found in the small amount of information available at the time. Although twenty years is quite a long period, it is not likely to be informative since we have the same information the all the database in each year. Thus, a priori, it is possible that the variables analyzed are only giving evidence of yearly effects. Nonetheless, the results are plausible and robust to changes in instruments, and to the exclusion of one or both of the other variables.

Table 6: IV-Probit	Set A	Set B	Set C
N° of Workers	0.188 ***	0.187 ***	0.186 ***
	0.012	0.012	0.012
Age	-0.028 ***	-0.031 ***	-0.033 ***
	0.007	0.007	0.007
Age Squared	0.000	0.000	0.000
	0.000	0.000	0.000
Sexratio	0.348 ***	0.352 ***	0.360 ***
	0.079	0.078	0.076
Educ.	0.103 **	0.094 *	0.070 *
	0.037	0.036	0.031
Educ. Squared	-0.002 *	-0.002	-0.002
	0.001	0.001	0.001
Educ*Wealth	-0.019 **	-0.019 **	-0.017 **
	0.007	0.007	0.006
Wealth	0.982 ***	1.108 ***	1.185 ***
	0.197	0.217	0.159
Wealth Squared	-0.056 *	-0.069 **	-0.087 ***
	0.025	0.026	0.018
Hist. Migration	0.087 ***	0.097 ***	0.104 ***
	0.010	0.012	0.006
Current Net	0.001	0.003	0.005 ***
	0.002	0.003	0.001
US Res.	0.076 ***	0.079 ***	0.081 ***
	0.013	0.013	0.012
Migration Prev.	11.446 **	9.938 **	6.433 **
	3.610	3.440	2.045
Mig. Prev.*Wealth	-1.466 **	-1.509 ***	-1.118 **
	0.514	0.423	0.372
Unem. Diff.	-0.876	-3.672	-5.221 *
	3.288	3.801	2.309
Exchange Rate	0.018	-0.004	-0.022
	0.023	0.028	0.012
Avg. Law index	-0.126 **	-0.091	-0.059 *
	0.042	0.047	0.025
Distance	0.001	0.000	0.000
	0.000	0.001	0.000
Border	0.209	-0.097	-0.365 *
	0.329	0.395	0.151
Constant	-6.553	-5.420	-4.030
	1.510	1.666	0.747
Wald test of exogeneity:			
Prob > chi2 =	0.022	0.0089	0.0582
Amemiya-Lee-Newey min. chi-sq statistic			7.902
P-value =			0.0952

Table 6 IV Probit. *, **, ***, stand for significance at 10%, 5% and 1% respectively

3.3 Tackling Simultaneously Self-Selection and Endogeneity: Three-Stage Estimation

Following Mroz (1987) and Piacentini (2008), a model able to tackle sample selection and endogeneity simultaneously has this reduced form:

$$l_i = \alpha^{lX_i} + \beta^{lN_i} + \varepsilon_{i1} \quad (11)$$

$$N_i = \theta Z_i + u_i \quad (12)$$

$$M_i = (\gamma^{R_i} + \varepsilon_{i2} > 0) \quad (13)$$

where equation (11) is level of migration, equation (12) identifies the instrument set, and equation (13) is selection. Errors are allowed to be arbitrarily correlated among the three. The system can be estimated through a TSLS, if at least two valid instruments are available.

Implementation consists of deriving the IMR from the first step of the Heckman correction method, and then using it as a regressor in the TSLS. Since *migration prevalence* is likely to be endogenous both in migration decision and in level of migration the three-step procedure is implemented as follow. The first step consists of estimating the probability of migration with an IV-Probit procedure. The second step consists of computing the IMR from the first step. The third step consists of using the IMR in the IV regression for level of migration.

Table 6 reports results for the level of migration equation, using all three sets of instruments. *Distance* and *border* are used as identification variables, and are therefore not included in the level of migration equation. This procedure was developed to study situations with one endogenous variable, and it needs at least two valid instruments. Thus, we rely on Set C for the discussion, since Set C has at least two valid instruments for each endogenous variable. However set A and set B tell us that even two instruments and two endogenous variables can produce consistent estimations.

As expected, and already observed, with a conventional HTS procedure, only a few variables have a significant influence on the optimal level of migration. In particular, human capital and household-level network variables all affect the number of migrants, as well as the probability of migration.

In the opposite direction, *unemployment differences* and *exchange rate* significantly affect the number of migrants, but not the migration decision. This is in line with picture (1b) and (1c). Households constrained in their optimal strategy by their budget are less likely to be affected by non-dramatic changes in the economic situation, since they cannot modify their migration strategy. In fact, they are likely to be non-migrants, or only one member migrates. Instead, richer households, able to send more members, are those more affected by changes in the economic situation.

The non-significance of community networks in the optimal number of migrant analysis confirms that community-level and household networks partially act as substitutes. Since they both affect the probability of migration, it is possible to argue that they have different functions, and/or that they convey different kinds of information.

Focusing on set A and set B we also observe a slightly significant inverted U-shaped relation between number of migrants and wealth. Once the problem of endogeneity is solved, IMR appears to be non-significant under sets A and B, and has a small negative effect under set C. This suggests, an absence of selectivity or, in the case of set C, negative bias selectivity, so that households with more “migration-prone” characteristics are those less likely to send more members.

Table 7: 2SLS Pr.	Set A	Set B	Set C
N° of Workers	0,268 **	0,082 ***	0,065 ***
	0,103	0,014	0,012
Age	-0,034 *	-0,010 ***	-0,008 ***
	0,014	0,002	0,002
Age Squared	0,000 ***	0,000 ***	0,000 ***
	0,000	0,000	0,000
Sexratio	0,440 *	0,073 *	0,039
	0,206	0,035	0,030
Educ.	0,138	0,008	-0,007
	0,074	0,010	0,009
Educ. Squared	-0,002	0,000	0,000
	0,001	0,000	0,000
Educ*Wealth	-0,029 *	-0,004	-0,001
	0,015	0,002	0,002
Wealth	1,173 *	0,181 *	0,073
	0,566	0,077	0,063
Wealth Squared	-0,066 *	-0,011 *	-0,007
	0,031	0,006	0,005
Hist. Migration	0,114 ***	0,056 ***	0,050 ***
	0,033	0,005	0,005
Current Net	0,004 ***	0,003 ***	0,003 ***
	0,001	0,001	0,001
US Res.	0,097 *	0,022 **	0,015 *
	0,043	0,008	0,007
Migration Prev.	12,132	1,004	-0,523
	6,529	0,816	0,667
Mig. Prev.*Wealth	-1,653	-0,204	0,041
	0,887	0,124	0,108
Unem. Diff.	-2,762 ***	-2,421 ***	-2,085 ***
	0,824	0,628	0,609
Exchange Rate	0,003	-0,014 ***	-0,015 ***
	0,009	0,002	0,002
Avg. Law index	-0,172 *	-0,021	-0,007
	0,086	0,014	0,012
Inverse Mills' Ration	1,208	-0,035	-0,150 *
	0,692	0,089	0,071
_cons	-6,532	-0,140	0,528
	3,614	0,463	0,371
Number of Obs.	12800	12800	12800
R2	0,123	0,228	0,230
Migration Prev.			
R-sq	0,848	0,857	0,8617
Adj R-sq	0,848	0,8567	0,8614
Shea's Partial R-sq	0,003	0,104	0,1738
Shea's Adj. P. R-sq	0,002	0,0977	0,1725
Robust F(2,12785)	152,830 ***	739,672 ***	324,644 ***
Mig. Prev*Wealth			
R-sq	0,759	0,7684	0,7731
Adj R-sq	0,759	0,7681	0,7727
Shea's Partial R-sq	0,004	0,0977	0,1508
Shea's Adj. P. R-sq	0,002	0,0965	0,1494
Robust F(2,12785)	183,048 ***	535,676 ***	211,048 ***
Overid. Test			8,261 0,08

Table 7 Second Stage of TSL IV. *, **, *** stand for significance at 0.10, 0.05 and 0.01, respectively

3.4 Alternative Econometric Specification: IV-Poisson

Since the level of migration is a count variable, TSL-IV regression may be inappropriate. This is the rationale behind the decision to propose the IV-Poisson as alternative specification. This is not the place to discuss the theoretical background behind IV-Poisson, the focus is on empirical results. A detailed explanation of the theoretical background is provided in Cameron and Trivedi (1998). In any case an important observation is necessary: IV-Poisson, assuming that the probability of each subsequent event is the same, cannot account for the selection bias.

This is in contrast with the underlying mechanism presumed at the beginning of this section. IV-Poisson assumes that there is no structural difference in sending one migrant or two. In economic terms, if there is a fixed cost of migration, it must be paid for each member undertaking migration, and not only for the first migrant. This is not implausible, recalling that an household is defined as migrant if one or more members have been in the U.S. in the three years prior to the undertaking of the survey, and that first-time migrants and experienced migrants are analyzed together. Therefore migration costs are plausibly different. In particular, the first migration requires an investment in contacts and human capital, while subsequent migrations probably do not.

Although with the limitation described, the count approach is technically more correct, as the dependent variable in question is a discrete number between 0 and 7. Table 7 reports results for Poisson, Negative Binomial and IV-Poisson regression.³⁸

The results do not contradict previous findings. Wealth affects migration with an inverted U-shaped relation, all household-level network variables positively affect the number of migrants. *Education* is non-significant, and no selection in terms of education is observed.

Migration prevalence and its cross-effect with *wealth* have the expected signs, but they are significant at the 0.05 confidence level only under set A. This reflects results already found in previous analyses. Migration has a certain fixed initial investment which affects the probability of migration (as found in IV-Probit analysis), but it does not affect the optimal number of migrants. We can presume that this initial investment cost is mitigated by community level networks. Since IV-Poisson estimates the number of migrants without a selection process, *migration prevalence* loses part of its significance. Networks of migrants are more important for poorer than for richer social strata, as highlighted by the negative coefficient of the cross-effect of community networks and wealth.

Coefficients for economic and policy variables have the expected signs, but they are non-significant.

³⁸ IV-Poisson regression is made with `ivpois` command in STATA10.

Table 8: IV-Poisson	Poisson	Neg. Binomial	Set A	Set B	Set C
N° of Workers	0.306 ***	0.326 ***	0.416 ***	0.375 ***	0.386 ***
	0.015	0.016	0.040	0.035	0.031
Age	-0.032 ***	-0.039 ***	-0.022	-0.033	-0.042 **
	0.009	0.010	0.020	0.020	0.016
Age Squared	0.000	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	0.000	0.000
Sexratio	0.639 ***	0.666 ***	0.752 **	0.557 **	0.558 **
	0.115	0.114	0.251	0.186	0.185
Educ.	0.089 *	0.081	0.301 *	0.104	0.122
	0.044	0.044	0.125	0.092	0.095
Educ. Squared	-0.004 *	-0.003 *	-0.003	-0.001	-0.002
	0.002	0.002	0.003	0.003	0.003
Educ*Wealth	-0.019 *	-0.019 *	-0.061 *	-0.032	-0.033
	0.009	0.009	0.025	0.020	0.019
Wealth	1.653 ***	1.781 ***	2.779 ***	2.329 ***	2.242 ***
	0.229	0.230	0.507	0.420	0.435
Wealth Squared	-0.140 ***	-0.155 ***	-0.136	-0.184 ***	-0.168 ***
	0.025	0.025	0.072	0.039	0.039
Hist. Migration	0.066 ***	0.083 ***	0.150 ***	0.182 ***	0.169 ***
	0.003	0.004	0.029	0.015	0.014
Current Net	0.006 ***	0.008 ***	0.012 **	0.014 ***	0.013 ***
	0.001	0.001	0.004	0.003	0.003
US Res.	0.084 ***	0.094 ***	0.121 ***	0.141 ***	0.137 ***
	0.015	0.015	0.032	0.026	0.026
Migration Prev.	2.277 **	2.514 ***	32.743 *	4.590	8.199
	0.729	0.750	13.910	4.989	5.559
Mig. Prev.*Wealth	-0.309 *	-0.329 *	-5.611 *	-1.322	-1.651
	0.147	0.151	2.341	0.804	0.926
Unem. Diff.	-10.368 ***	-9.162 **	-9.097	-17.859	-11.561
	2.940	2.895	7.735	11.044	6.219
Exchange Rate	-0.064 ***	-0.058 ***	-0.027	-0.095	-0.063 *
	0.010	0.011	0.043	0.052	0.029
Avg. Law index	-0.131 ***	-0.121 ***	-0.078	-0.069	-0.078
	0.032	0.031	0.068	0.058	0.046
_cons	-4.547	-5.217	-13.734	-5.043	-6.533
	0.691	0.683	3.417	2.639	2.332
Log-Likelihood	-6203.355	-6112.007			
Test of Exogeneity					
Migration Prev.			***	***	***
Mig. Prev.*Wealth			***	***	***

Table 8. IV-Poisson . *, **, ***, stand for significance at 0.10, 0.05 and 0.01, respectively

4 Discussion and Conclusions

The main contribution of this paper is to reassess the analysis on the determinant of migration by focusing not only on the probability of migration, rather on the optimal investment in migration. To do so an empirical approach new to international migration literature was applied to evidence the relation between migration choices, wealth, costs and migrant networks.

In line with previous findings, Mexican migrants belong to the middle of the income distribution in Mexico. Migration and wealth are non-linearly related. Household and community networks increase the migration propensity. Specifically, when large enough, networks further increase the migration propensity of the households belonging to the middle-left of the income distribution.

Community and household-level networks are partially substitutes and partially complements. In particular, while household-level networks always positively affect migration (in both terms of propensity and numbers), community-level networks convey information which makes migration a feasible strategy.

The Three-Stage Procedure and the IV-Poisson can simultaneously solve several empirical problems typical of migration studies: sample selection, endogeneity of migration networks, and the presence of count dependent variables. These approaches confirm previous findings, ensuring that they are not the result of endogeneity or sample-selection. Nonetheless, both methods, and more in general empirical migration studies, need to be improved.

Particular promising seems to be the analysis of the legislation. The effect of U.S. legislation on Mexican migration should be investigated in more detail, focusing not on the aggregate level of analysis, but on checking whether changes in legislation have generated changes in the composition of migration flows. Since MMP contains information on the first and last destinations of household heads in the U.S. and Canada, local changes in legislation could be collected and analyzed to see if they affect the destinations of migrants.

References

- Ai, C., & Norton, E. C. (2003, July). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123-129.
- Borjas, G. J. (1985, Oct.). Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics*, 77(4), 463-89.
- Borjas, G. J. (1987, Sep.). Self-Selection and the Earnings of Immigrants. *The American Economic Review*, 77(4), 531-553.
- Borjas, G. J. (1990, Mar.). Self-Selection and the Earnings of Immigrants: Reply. *The American Economic Review*, 80(1), 305-308.

- Borjas, G. J. (1991, May). Immigrants in the U.S. Labor Market: 1940-80. *The American Economic Review, Papers and Proceedings of the Hundred and Third Annual Meeting of the American Economic Association*, 81(2), 287-291.
- Borjas, G. J. (1994, Dec.). The Economics of Migration. *Journal of Economic Literature*, 32(4), 1667-1717.
- Borjas, G. J. (1995, Apr.). Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s? *Journal of Labor Economics*, 13(2), 201-245.
- Borjas, G. J. (1995, Spring). The Economic Benefits from Immigration. *The Journal of Economic Perspectives*, 9(2).
- Cameron, A. C., & Trivedi, P. K. (1998). *Regression Analysis of Count Data* (Vol. Econometric Society Monograph No.30). Cambridge University Press.
- Chiswick, B. R. (1999, May). Are Immigrants Favorably Self-Selected? *The American Economic Review*, 89(2), 181-185.
- Chiswick, B. R., & Miller, P. W. (2002). Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal of Population Economics*, 15, 31-57.
- Chiswick, B. R., & Miller, P. W. (2006). Computer usage, destination language proficiency and the earnings of natives and immigrants. *KMI Working Paper Series*.
- Clemens, M. A., Montenegro, C. E., & Pritchett, L. (2008, July). The Place Premium: Wage Differences for Identical Workers across the US Border. *Background Paper to the 2009 World Development Report*(4671).
- Cornelius, W. (2005). Controlling 'Unwanted' Immigration: Lessons from the United States, 1993-2004. *Journal of Ethnic and Migration Studies*, 4(31), 775-794.
- de Haan, A. (2006, Feb.). Migration in the Development Studies Literature. *UNU-WIDER Research Paper*(2006/19).
- Docquier, F., & Rapoport, H. (2003). Remittances and Inequality: A Dynamic Migration Model. *IZA Discussion Papers*(808).
- Durand, J., Massey, D. S., & Zenteno, R. M. (2001). Mexican Immigration to the United States: Continuities and Changes. *Latin American Research Review*, 36(1), 107-127.
- Fawcett, J. T. (1989). Networks, Linkages, and Migration Systems. *International Migration Review*, 23(3), 671-680.
- Filmer, D., & Pritchett, L. H. (2001, Feb.). Estimating Wealth Effect without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. *Demography*, 38(1), 115-132.
- Fussell, E., & Massey, D. S. (2004, Feb.). The Limits to Cumulative Causation: International Migration from Mexican Urban Areas. *Demography*, 41(1), 151-171.

- Ghatak, S., Levine, P., & Wheatley Price, S. (1996). Migration Theories and Evidence: an Assessment. *Journal of Economic Surveys*, 10(2), 159-198.
- Goyal, S. (2007). *Connections An introduction to Economics of Networks*. Princeton, New Jersey: Princeton University Press.
- Hanson, G. H. (2006, December). Illegal Migration from Mexico to United States. *Journal of Economic Literature*, 94, 869-924.
- Hanson, G. H., & Spilimebrgo, A. (1999, Dec.). Illegal Immigration, Border Enforcement, and Relative Wage: Evidence from Apprehensions at the U.S.-Mexico Border. *The American Economic Review*, 89(5), 1337-1357.
- Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two sector analysis. *America Economic Review*, 60, 126-142.
- Hatton, T. J., & Williamson, J. G. (1992, September). International Migration and World Development: a Historical Perspective. *NBER Historical Paper*(41), 1-79.
- Hatton, T. J., & Williamson, J. G. (2003, March). What Fundamentals Drive World Migration? *World Institute for Development Economics Research Discussion Paper*(2003/23).
- Konseiga, A. (2006). Household Migration Decisions as Survival Strategy: The Case of Burkina Faso. *Journal of African Economies* , 16(2), 198-233.
- Massey, D. S., & Espinosa, K. E. (1997, January). What's driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis . *The American Journal of Sociology*, 102(4), 939-999.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993, September). Theories of International Migration: A Review and Appraisal. *Population and Development Review*, 19(3), 433-466.
- Massey, D. S., Goldring, L., & Durand, J. (1994, May). Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities. *The American Journal of Sociology*, 99(6), 1492-1533 .
- McKenzie, D. (2005). Measuring inequality with asset indicators. *Journal of Population Economics*, 18, 229-260.
- McKenzie, D., & Rapoport, H. (2007). Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. *Journal of Development Economics*, 84, 1-24.
- Mexican Migration Project. (s.d.). <http://mmp.opr.princeton.edu/>.
- Mroz, T. A. (1987, Jul.). The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica*, 55(4), 765-799.
- Newey, W. K., & West, K. D. (1987, May). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent. *Econometrica*, 55(3), 703-708.

- Orrenius, P. M. (1999, December). The role of family networks, coyote prices and the rural economy in migration from Western Mexico: 1965-1994. *Federal Reserve Bank of Dallas, Working Paper, 9910*, 1-44.
- Orrenius, P. M. (2001). Illegal immigration and enforcement along the southwest border. *Border Economy*(8).
- Ortega, F., & Peri, G. (2009). The Causes and Effects of International Labor Mobility: Evidence from OECD Countries 1980-2005. *NBER Working Paper Version*(14833).
- Peri, G. *Giovanni Peri website*. <http://www.econ.ucdavis.edu/faculty/gperi/>
- Piacentini, M. (2008). Migration Enclaves, Schooling Choices and Social Mobility. *Centro Studi Luca d'Agliano Development Working Papers*.
- Radu, D. (2008, May 531-548). Social Interactions in Economic Models of Migration: A Review and Appraisal. *Journal of Ethnic and Migration Studies* , 34(4).
- Roy, A. D. (1951, Jun.). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers, New Series*, 3(2), 135-146.
- Sjaastad, L. A. (1962, Oct.). The Costs and Returns of Human Migration. *The Journal of Political Economy*, 70(5), 80-93.
- Stark, O. (1984). Migration decision making: A review article. *Journal of Development Economics*, 14, 251-259.
- Stark, O. (1991). *The migration of Labor*. Cambridge: Basil Blackwell.
- Stark, O., & Levhari, D. (1982, Oct.). On Migration and Risk in LDCs. *Economic Development and Cultural Change*, 31(1), 191-196.
- Taylor, J. E. (1986). Differential migration, networks, information and risk. In O. Stark, *Research in Human Capital and Development* (Vol. 4, p. 147-171). Greenwich, Connecticut, US: JAI Press.
- Todaro, M. (1969). A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, 59, 138-148.
- Winters, P., de Janvry, A., & Sadoulet, E. (2001, Winter). Family and Community Networks in Mexico-U.S. Migration. *The Journal of Human Resources*, 36(1), 159-184.
- Woodruff, C., & Zenteno, R. (2007). Migration networks and microenterprises in Mexico. *Journal of Development Economics*, 82, 509-528.
- Yang, D. (2008, April). International Migration, Remittances and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks. *The Economic Journal*(118), 591-630.

Appendix OLS

Probability of Migration	OLS	OLS Rob. SE	OLS Cluster Rob. SE
N° of Workers	0.038 *** 0.002	0.038 *** 0.003	0.038 *** 0.004
Age	-0.009 *** 0.001	-0.009 *** 0.001	-0.009 *** 0.001
Age Squared	0.000 *** 0.000	0.000 *** 0.000	0.000 *** 0.000
Sexration	0.053 *** 0.014	0.053 *** 0.013	0.053 *** 0.015
Educ.	0.002 0.005	0.002 0.004	0.002 0.004
Educ. Squared	0.000 0.000	0.000 0.000	0.000 0.000
Educ*Wealth	-0.001 0.001	-0.001 0.001	-0.001 0.001
Wealth	0.156 *** 0.022	0.156 *** 0.018	0.156 *** 0.025
Wealth Squared	-0.014 *** 0.003	-0.014 *** 0.002	-0.014 *** 0.003
Hist. Migration	0.030 *** 0.001	0.030 *** 0.002	0.030 *** 0.004
Current Net	0.001 *** 0.000	0.001 *** 0.000	0.001 ** 0.000
US Res.	0.020 *** 0.003	0.020 *** 0.003	0.020 *** 0.004
Migration Prev.	0.367 *** 0.107	0.367 *** 0.104	0.367 0.248
Mig. Prev.*Wealth	-0.049 * 0.021	-0.049 * 0.021	-0.049 0.039
Unem. Diff.	-0.856 * 0.357	-0.856 * 0.357	-0.856 0.813
Exchange Rate	-0.005 *** 0.001	-0.005 *** 0.001	-0.005 0.003
Avg. Law index	-0.008 0.004	-0.008 0.005	-0.008 0.012
Distance	0.000 0.000	0.000 0.000	0.000 0.000
Border	-0.061 *** 0.015	-0.061 *** 0.013	-0.061 0.036
Constant	-0.021 0.071	-0.021 0.060	-0.021 0.125
N° of observations	12805	12805	12805
R2	0.1933	0.1933	0.1933

Table 9: OLS. *, **, ***, stand for significance at 0.10, 0.05 and 0.01, respectively

Number of Migrants	OLS		OLS	
			Rob. SE	Cluster Rob. SE
N° of Workers	0.087 ***	0.087 ***	0.087 ***	0.087 ***
	0.004	0.006	0.013	
Age	-0.011 ***	-0.011 ***	-0.011 ***	-0.011 ***
	0.002	0.002	0.002	
Age Squared	0.000 ***	0.000 ***	0.000 **	0.000 **
	0.000	0.000	0.000	
Sexration	0.084 ***	0.084 ***	0.084 **	0.084 **
	0.022	0.021	0.025	
Educ.	0.005	0.005	0.005	
	0.008	0.006	0.006	
Educ. Squared	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	
Educ*Wealth	-0.003	-0.003	-0.003	-0.003
	0.002	0.001	0.001	
Wealth	0.177 ***	0.177 ***	0.177 ***	0.177 ***
	0.034	0.027	0.039	
Wealth Squared	-0.015 ***	-0.015 ***	-0.015 ***	-0.015 ***
	0.004	0.003	0.004	
Hist. Migration	0.056 ***	0.056 ***	0.056 ***	0.056 ***
	0.001	0.004	0.004	
Current Net	0.003 ***	0.003 ***	0.003 **	0.003 **
	0.000	0.001	0.001	
US Res.	0.024 ***	0.024 ***	0.024 **	0.024 **
	0.004	0.006	0.009	
Migration Prev.	0.212	0.212	0.212	
	0.169	0.160	0.309	
Mig. Prev.*Wealth	-0.005	-0.005	-0.005	-0.005
	0.034	0.034	0.059	
Unem. Diff.	-1.978 ***	-1.978 ***	-1.978	-1.978
	0.564	0.576	1.325	
Exchange Rate	-0.010 ***	-0.010 ***	-0.010 *	-0.010 *
	0.002	0.002	0.004	
Avg. Law index	-0.023 ***	-0.023 **	-0.023	-0.023
	0.007	0.008	0.018	
Distance	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	
Border	-0.062 **	-0.062 **	-0.062	-0.062
	0.024	0.019	0.051	
Constant	-0.125	-0.125	-0.125	-0.125
	0.111	0.089	0.187	
N° of observations	12805	12805	12805	12805
R2	0.2325	0.2325	0.2325	0.2325

Table 10: OLS (Number of Migrants) *, **, ***, stand for significance at 0.10, 0.05 and 0.01, respectively

HTSP	Selection		Selection		Numb.	Numb.	
		Rob. SE	Cluster	Rob. SE		Rob. SE	Cl. Rob. SE
N° of Workers	0.186 ***	0.186 ***	0.186 ***	0.245 ***	0.245 ***	0.245 ***	
	0.012	0.013	0.013	0.020	0.030	0.034	
Age	-0.033 ***	-0.033 ***	-0.033 ***	-0.013	-0.013	-0.013	
	0.007	0.007	0.007	0.008	0.007	0.007	
Age Squared	0.000	0.000	0.000	0.000	0.000	0.000	
	0.000	0.000	0.000	0.000	0.000	0.000	
Sexratio	0.363 ***	0.363 ***	0.363 ***	0.351 ***	0.351 ***	0.351 **	
	0.076	0.082	0.084	0.092	0.097	0.117	
Educ.	0.042	0.042	0.042	0.056	0.056	0.056	
	0.028	0.030	0.031	0.036	0.031	0.033	
Educ. Squared	-0.002 *	-0.002	-0.002 *	0.001	0.001	0.001	
	0.001	0.001	0.001	0.001	0.001	0.001	
Educ*Wealth	-0.011	-0.011	-0.011	-0.015 *	-0.015 *	-0.015 *	
	0.006	0.007	0.006	0.007	0.006	0.007	
Wealth	1.112 ***	1.112 ***	1.112 ***	0.461 *	0.461 *	0.461 **	
	0.144	0.180	0.174	0.209	0.211	0.177	
Wealth Squared	-0.101 ***	-0.101 ***	-0.101 ***	-0.031	-0.031	-0.031	
	0.016	0.019	0.018	0.021	0.020	0.018	
Hist. Migration	0.103 ***	0.103 ***	0.103 ***	0.049 ***	0.049 ***	0.049 ***	
	0.004	0.007	0.015	0.007	0.012	0.007	
Current Net	0.005 ***	0.005 ***	0.005 **	0.005 ***	0.005 ***	0.005 ***	
	0.001	0.001	0.002	0.001	0.002	0.001	
US Res.	0.081 ***	0.081 ***	0.081 ***	0.033 *	0.033 *	0.033 *	
	0.012	0.010	0.013	0.013	0.016	0.014	
Migration Prev.	1.986 ***	1.986 ***	1.986	0.331	0.331	0.331	
	0.568	0.521	1.445	0.790	0.816	0.872	
Mig. Prev.*Wealth	-0.278 *	-0.278 *	-0.278	0.033	0.033	0.033	
	0.114	0.112	0.228	0.150	0.155	0.164	
Unem. Diff.	-3.853 *	-3.853 *	-3.853	-5.262 *	-5.262 *	-5.262	
	1.834	1.765	4.026	2.116	2.532	3.617	
Exchange Rate	-0.022 **	-0.022 **	-0.022	-0.034 ***	-0.034 ***	-0.034 **	
	0.007	0.007	0.015	0.009	0.009	0.012	
Avg. Law index	-0.045 *	-0.045 *	-0.045	-0.093 ***	-0.093 ***	-0.093 **	
	0.020	0.019	0.048	0.021	0.023	0.030	
Distance	0.000	0.000	0.000				
	0.000	0.000	0.000				
Border	-0.393 ***	-0.393 ***	-0.393 ***				
	0.086	0.078	0.239				
Constant	-3.082	-3.082	-3.082	-1.247	-1.247	-1.247	
	0.437	0.462	0.837	0.670	0.772	0.656	
Inverse Mills ratio				0.510 ***	0.510 **	0.510 ***	
				0.127	0.172	0.148	
Censored obs				10781	10781	10781	
Uncensored obs				2024	2024	2024	

Table 4 HTSp. *, **, ***, stand for significance at 0.10, 0.05 and 0.01, respectively. Robust standard errors using bootstrapping method



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
<http://www.dse.unibo.it>