COMPUTATIONAL ECONOMIC MODELING OF MIGRATION

Anna Klabunde

Ruhr University Bochum, Germany

Universitaetsstr. 150

44780 Bochum

anna.klabunde@rub.de

COMPUTATIONAL ECONOMIC MODELING OF MIGRATION

ABSTRACT

In this chapter an agent-based model of endogenously evolving migrant networks is developed to find and estimate the size of determinants of migration and return decisions. Individuals are connected by links, the strength of which declines over time and distance. Methodologically, this paper combines parameterization using data from the Mexican Migration Project with calibration. It is shown that expected earnings, an idiosyncratic home bias, network ties to other migrants, strength of links to the home country and age have a significant impact on circular migration patterns over time. The model can reproduce spatial patterns of migration as well as the distribution of number of trips of migrants. It can also be used for computational experiments and policy analysis.

JEL-Classification: C63; F22; J61

Keywords: Circular migration, social networks, agent-based computational economics

1 Introduction

59 % of Mexican migrants to the US surveyed in the Mexican Migration Project (MMP128) make more than one move, i.e., after returning to Mexico they go back to the US at least once. The phenomenon makes it difficult to forecast stocks of migrants in the US at any point in time and to make estimates of where they are likely to go and when, if at all, they are going to return. So far, research on so-called circular migration has mostly been empirical, using multinomial logit, count data models, duration models or Markov transition matrices to estimate migration and return probabilities controlling for individual and/or home or host country characteristics. Examples are Constant and Zimmermann (2003 and 2011), Bijwaard (2010), Vadean and Piracha (2009) and Reyes (2001). A detailed empirical study on motives of Mexican circular migrants, testing different theories using the MMP, is Massey and Espinosa (1997). Hill (1987) is an attempt at formalizing duration of stay and frequency of trips in a life-cycle model. A more recent theoretical model of circular migration is Vergalli (2011), who studies the phenomenon in a real option framework, with the duration of stay modeled as hysteresis between thresholds that trigger entry and exit.

When developing a model that is sufficiently realistic to be used for policy analysis or, eventually, forecasts and that is empirically founded one has to take into account some important aspects of the issue at hand:

First, a migrant's decision is not independent from that of other migrants and potential migrants. Other migrants support the newly arrived in their job search, and home-community members help return migrants to reintegrate into the home country labor market. The role of social networks in migration decisions has been subject of substantial research; Radu (2008) provides a survey of the literature. It is often observed that: 1. Migration concentrates on a certain number of places. 2. People from one neighborhood tend to go to the same few places. 3. Migration specific capital makes subsequent migrations more likely.

Second, since a migrant expands his or her network with every migration move and network ties possibly become weaker over time, different parts of the migration cycle should not be seen separately.

Third, an individual's decision to move creates two externalities: one on the network at the location of origin, and the other one on the network in the destination country. If one migrant leaves, others might leave as well. Whether or not the migrant returns, the size of the destination country network will have changed through his or her move. Thus when he or she considers migrating again, the conditions have changed compared to the previous move, partly caused by his or her own behavior. Hence, there is happening a repercussion process, with the network influencing the migrant, the migrant influencing the network, and the new network influencing the migrant. This has been dubbed the "reflection problem" by Manski (1993). How large is the effect of networks on both migration and return decision, and what other possible determinants of circular migration exist?

In order to approach this question and to create a space for policy experiments having to do with (circular) migration I build an agent-based model that allows for the necessary modeling flexibility and for the spatial dimension of the problem. Its central component is the role of networks that evolve endogenously from migration decisions. Links decay over time and physical distance. The migration behavior of one generation of heads of household is modeled over a period of 33 years.

There are some rather simple, uncalibrated agent-based models on migration (Makowsky *et al.* 2006, Silveira *et al.* 2006, Espíndola *et al.* 2006). The present model, in turn, is one of the few examples of completely calibrated and empirically founded agent-based models that deal with migration or networks. Related models include Feitosa (2010), who studies urban segregation in Brazil, Sun and Manson (2010) on housing search in Minnesota, Haase *et al.* (2010) and Fontaine and Rounsevell (2009) on residential mobility, and Mena *et al.* (2011) and Entwisle *et al.* (2008) who model land use change. The former includes migration behavior - although it is not the focus of the model - and the latter social networks. A very recent paper by Kniveton et al. (2011) replicates climate-induced regional migration flows in Burkina Faso using an agent-based model with networks as information transmission mechanism. A different computational approach at empirically founded models of Mexican circular migration has been introduced in very recent papers in which discrete choice dynamic programming models are estimated using Maximum Likelihood (Lessem 2011) or the Simulated Method of Moments (Thom 2010).

In the present model, the MMP128 and other data sources were used for parameterization. The MMP is a large event-history microsurvey data set of Mexican migrants and non-migrants. The data set is described in Durand *et al.* (2001). Its potential shortcomings are outlined in Thom (2010).

Parameters that cannot be found easily in econometric models due to endogeneity problems and the spatial dimension are calibrated such that parameter values are found that create a close match between simulated and observed data.

Proceeding in this way a common criticism of agent-based-models is avoided, namely many degrees of freedom and the resulting possibility to create almost any desired output. All of the parameters except four are fixed. Those remaining four are calibrated indirectly by matching the simulated data to real data: The distribution of number of trips of migrants, the distribution of migrants across US cities, and the time series of percent of agents migrating and returning per year. It is then possible to perform experiments with the model.

The paper is structured as follows: Section 2 introduces three stylized facts on circular migration that the model should match. Section 3 derives and tests hypotheses on behavioral motives to include in the model. Section 4 describes the model, which is parameterized in Section 5. The indirect calibration procedure and results are described in Section 6. In Section 7 an example is provided on how to use the model for policy experiments. Section 8 concludes.

2 STYLIZED FACTS ON CIRCULAR MIGRATION

From the literature and the MMP128 three stylized facts on circular migration can be derived that will be attempted to match with the model. If the model succeeds in recreating these prominent characteristics of circular migration behavior it is a plausible candidate for the true data generating process.

2.1 Migration concentrates on a certain number of places

The bulk of migration originates in a few places and it concentrates on a fairly small number of places in the country of destination. In the case of Mexican migration to the US, the communities with the highest percentage of adults with migrant experience are in the states of Guanajuato, Durango, Jalisco and Michoacàn (MMP128). The percentage varies from just above one percent to almost 50 across communities. Of the migrating heads of household surveyed in the MMP128, 20% went to the Los Angeles district on their last trip; by far the highest number, followed by the Chicago region (8%) and the San Diego region (5%). Since 62% of migrants went to California on their last move I choose to model the migration behavior of one generation of Mexican heads of household who - if they chose to migrate - went from Western-central Mexico to California.

In order to calibrate the model to the empirical distribution and to have a means to validate the model, the distribution of migrants across cities is determined. When observing the complete MMP128 sample, the distribution across cities is very similar to the Western-Mexico California subsample. Therefore, both the subsample and the full sample are used in order to avoid bias in the estimates due to small sample size.

Distributions that result from social interaction often follow a power law; see Clauset *et al.* (2007), Axtell (2001) and Redner (1998) for examples. Indeed there is a small number of cities that attract a very large proportion of migrants, and many cities attract only one migrant. This indicates that the distribution might follow a power law.

To check whether the number of migrants in a city from the large sample is actually distributed according to a power law I follow the methodologies suggested by Goldstein *et al.* (2004) and Clauset *et al.* (2007).

I also apply the method and Matlab routines provided by Clauset *et al.* (2007) that include estimating a minimum value for x above which the power law applies. Maximum likelihood estimation is used for the exponent of the zeta distribution defined as

$$p(k) = \frac{k^{-\gamma}}{\zeta(\gamma, x_{min})} \tag{1}$$

where k is a positive integer measuring, in this case, number of migrants in a city, $p(k)$ is the probability of observing the value k, γ is the power law exponent, $\zeta(\gamma)$ is the Riemann zeta function defined as $\sum_{k=1}^{\infty} k^{-\gamma}$, and x_{min} is a minimum value for x above which the power law applies. This yields a minimum x value of 34 and a scale parameter of 1.9.

Although visually the power law seems to be a good fit (not shown), I check whether the distribution might actually follow a power law above the minimum x value of 34. The test statistic is the maximum distance between the cumulative distribution function of a power law distribution with exponent γ and the cumulative empirical distribution. The idea is to determine the proportion of random draws from the hypothesized distribution that yield a larger distance than the empirical distribution. The larger this proportion, the better is the fit between the empirical distribution and the hypothesized distribution (Clauset *et al.* 2007). A large number of synthetic power law distributions is generated, the parameter γ for each distribution is estimated and the Kolmogorov-Smirnov-statistic is computed. Then the proportion of those values is determined in which the KS-statistic is larger than the one from the empirical distribution. If the proportion p is such that $p > .1$, the power law hypothesis is not rejected. In the present case, $p = .4250$, so the power law hypothesis cannot be rejected.

I follow the same procedure for the smaller subsample that I use as basis for the simulation. The results indicate that even for the small subsample the distribution might follow a power law for values larger than 15 with $\gamma = 2.2$ (see fig. 1).

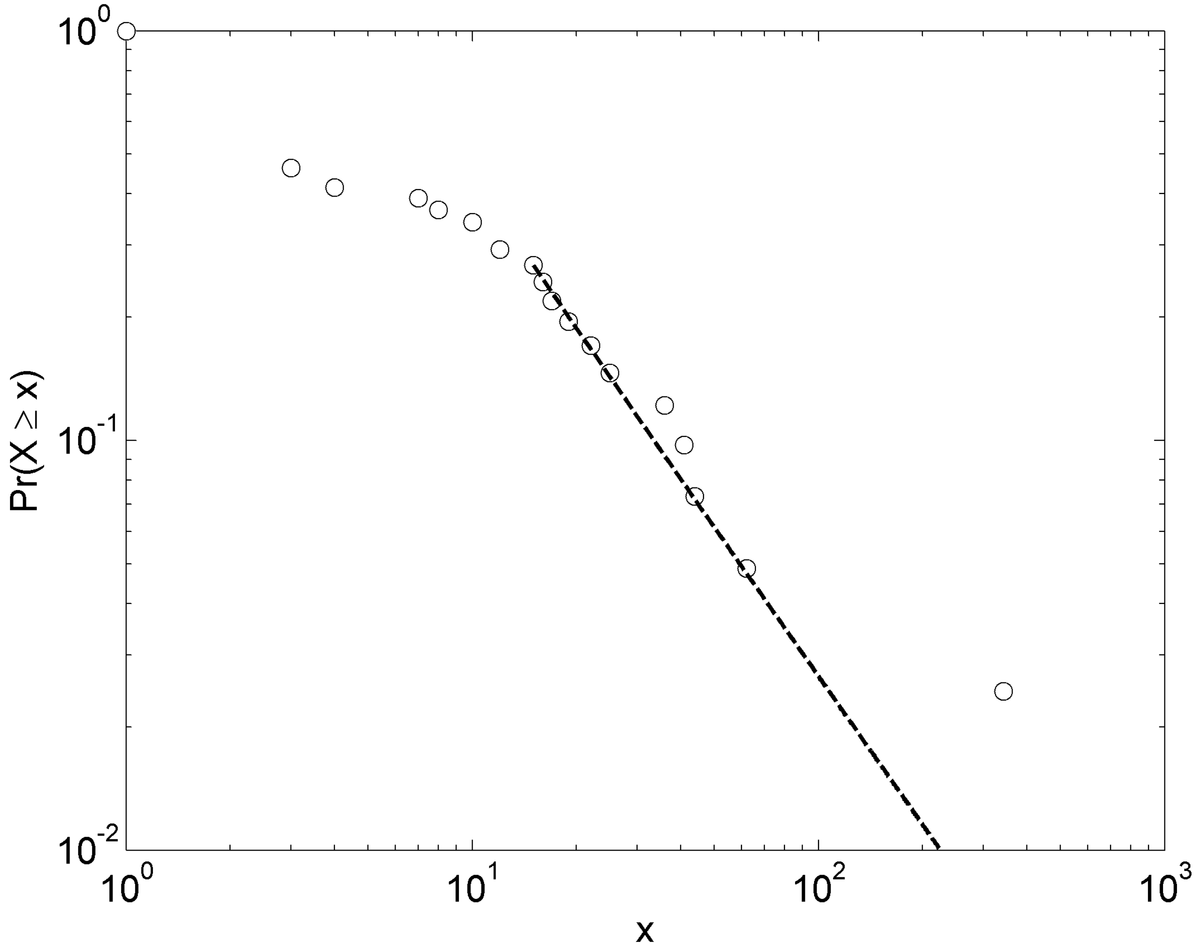

Figure 1: Log-log plot of the cumulative distribution function of numbers of migrants per city in the small subsample and fitted values using MLE, with $\gamma = 2.2$.

This hypothesis is not rejected in a Kolmogorov-Smirnov-test either, since the p-value of .845 is very high. However, these results have to be taken with caution due to the small sample size.

For comparing the empirical to the simulated distribution at the end of the calibration procedure, mean, standard deviation and median of the two distributions are compared. Whether the simulated distribution resembles a power law, both for single runs and for the overall distribution after 10,000 model runs, will be checked (see Section 6.2).

2.2 People from one neighborhood tend to go to the same few places

People tend to settle where people from the same region of origin have settled previously: 65% of the migrant heads of household surveyed in the community of the highest percentage of migrants among the population, a village in Michoacán, went to the Chicago region. Patterns in most other communities are very similar. For additional evidence see Munshi (2003) and Bauer *et al.* (2007). The reasons for this are positive network externalities.

2.3 Migration specific capital makes subsequent migration moves more likely

Several studies reveal the importance of migration specific capital, i.e. experience and knowledge that facilitate every subsequent move. This migration specific capital is closely related to migrant networks as well: with every move, migrants build up new links that facilitate job search, (re)integration and information flow (DaVanzo 1981). Therefore, once a move has taken place, migrants are more prone to move (again) than they were before their first move (Constant and Zimmermann 2011).

Since some of the individuals in the subsample were interviewed before the last year considered (2007) and therefore their migration histories are not complete, the full sample is used

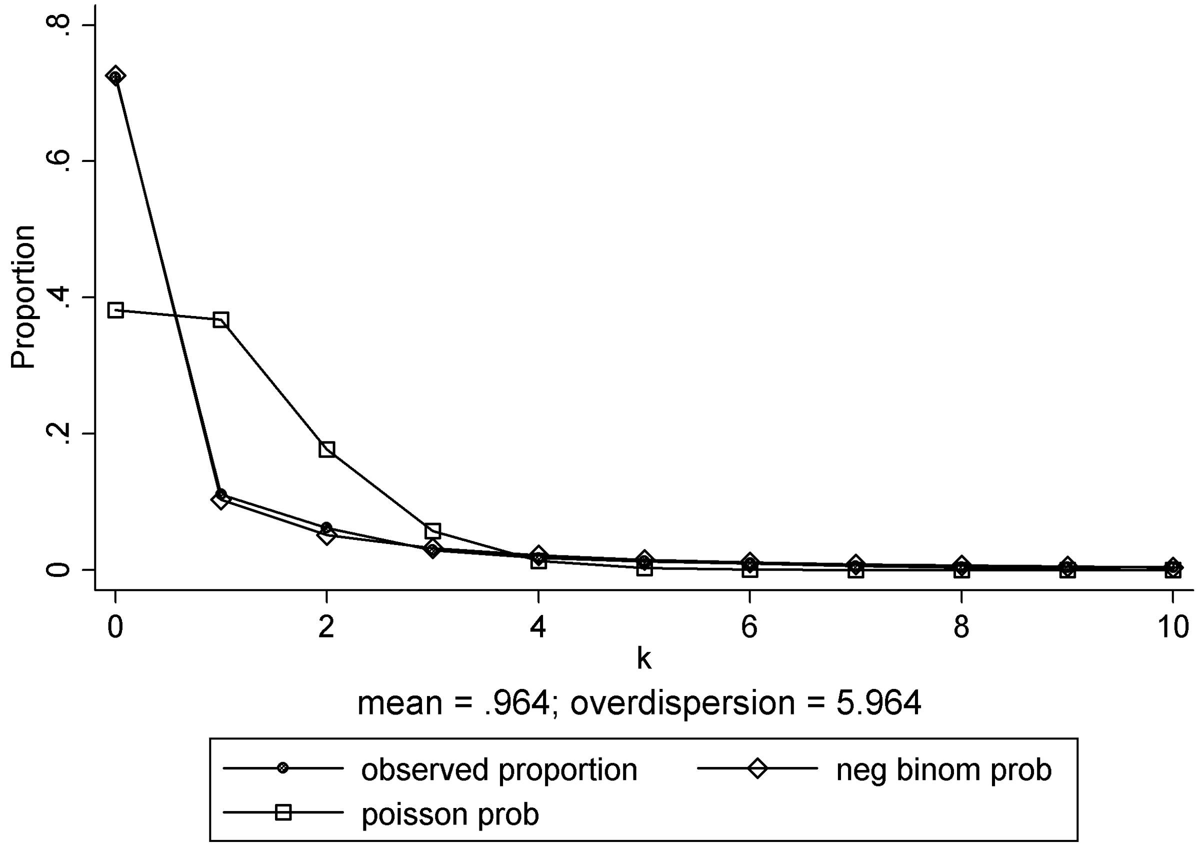

Figure 2: Observed distribution of number of trips compared to a Poisson and a negative binomial distribution.

for measuring the distribution of number of trips. The total number of trips is measured at age 47, which corresponds to the last year in the lives of the simulated agents. The distribution of number of trips displays overdispersion (mean = .964, standard deviation = 2.785) and "excess zeros" as compared to a Poisson distribution. The observed distribution fairly closely resembles a negative binomial one (see fig. 2). In fact, the null-hypothesis that it is equal to a negative binomial one could not be rejected in a Kolmogorov-Smirnov test. The overdispersion and "excess zeros" could be due to either heterogeneity of individuals, or to two different data generating mechanisms creating zero and nonzero counts of trips (Greene 2003, 744-752). Both explanations would be in line with the argument by DaVanzo (1981) and Constant and Zimmermann (2011): Migrants could have characteristics that distinguish them from non-migrants, so the heterogeneity between people who do not migrate at all (number of trips = 0) and people who do make one trip would be much larger than between migrants who make one trip and migrants who make two trips. Alternatively, the conditions for making the first trip are much different from those for subsequent trips due to the above mentioned build-up of migration specific capital. Therefore, the mechanism "generating" 0 moves is different from the one generating a positive number of trips.

The model developed here is useful if it succeeds in recreating these three stylized facts.

3 SELECTION OF BEHAVIORAL MOTIVES

Several behavioral motives can be found in the literature that might influence migration and/or return decision. Which ones of those to include in the model is determined by running logit and probit regressions on the MMP128 data set for the probability to migrate and to return in a person-year. The full sample of individuals for the years 1970-2008 was used, thereby implicitly assuming that they are not systematically different from the Western Mexican subsample which is used for simulation. All of the hypotheses mentioned below are included in the regressions, as well as controls for family status, community of origin, profession and current job. All results are displayed in table 1.

Hypothesis 1: Expected earnings

The first hypothesis is: Migrants are attracted by a higher expected income in the host country than in the home country, taking into account the unemployment rate (Harris and Todaro 1970). The higher the expected income as compared to the current income, the more likely someone is to migrate.

It is not straightforward to find the effect of the difference between expected earnings and current earnings on the migration and return decision with the data available from the MMP128, for two reasons: 1. The data does not contain information on earnings of every person-year, but only on the year of the survey and of the first and last migration. 2. It is

	Probit Prob. for trip	**1. Probit Prob. for return if in US**	**2. Logit Prob. for return if in US**
Variables	**Coefficients**		
Sex (female =1)	-.265***	-.313***	-.550***
	(.033)	(.064)	(.113)
Age 18 to 30 (reference: < 18)	.135***	.248***	.464***
	(.020)	(.051)	(.087)
Age 31 to 45	-.173***	.258***	.491***
	(.023)	(.055)	(.094)
Age 46 to 60	-.509***	.121*	.225**
	(.027)	(.064)	(.110)
> 60	-.998***	.168*	.231
	(.047)	(.095)	(.163)
No. family of origin ever migrated	.076***	-.025***	-.040***
	(.003)	(.004)	(.008)
Married	-.033***	.087***	.158***
	(.012)	(.022)	(.038)
Number of children	- .00001	.017***	.031***
	(.00005)	(.004)	(.007)
Property index categories (reference= 0)			
Property index = 1	-.049***		
	(.012)		
Property index = 2	-.109***		
	(.019)		
Property index = 3	-.171***		
	(.024)		
Property index = 4	-.106***		
	(.028)		
Green card		-.00004**	-.00006**
		(.00002)	(0.00003)
Documentation used for trip (reference "unknown")			
legal resident		.342**	.537*
		(.172)	(.303)
Contract - Bracero		.944***	1.680***
		(.352)	(.653)
Contract - H2A (agricultural)		.595***	.982***
		(.204)	(.358)
Temporary: Worker		.471**	.765**
		(.210)	(.362)
Temporary: Tourist		.739***	1.235***
		(.177)	(.311)
Citizen		.366*	.473
		(.207)	(.368)
Undocumented / false documents		.680***	1.101***
		(.171)	(.302)
Years since last trip	-.058***	-.077***	-.151***
	(.002)	(.003)	(.006)
Number of previous trips	.084***		
	(.002)		
Property index larger in t+1 (1 = yes)		-.599***	-1.093***
		(.039)	(.069)
Last US wage PPP (thousands)		-.0009***	-.002***
		(.00005)	(.00008)
Last US wage PPP (thousands) * Property index larger in t+1		.0009***	.0002***
		(.00009)	(.00002)
Exp. annual wage-difference US-Mexico (thousand USD)	.033***	-.005	-.013**
	(.001)	(.003)	(.006)
Before first migration? (yes =1)	-.875***		
	(.016)		
Constant	-2.907***	-.712**	-1.098*
	(.093)	(.317)	(.582)
Number of observations	510578	32709	32709
Pseudo R^2	0.426	0.292	0.296

Table 1: Probability of moving. Years 1970-2007. County and occupation dummies were used. Robust standard errors in parentheses; *** significant at least at the 1% level, **significant at least at the 5% level, * significant at least at the 10% level.

Number of previous trips	**Increase in probability to do a trip in person-year per 1,000 USD expected wage difference**	**z**	$P > \|z\|$
none	.0011 (.00005)	21.54	0.00
at least 1	.0035 (.0002)	21.99	0.00

Table 2: Predictive margins obtained via delta method. Standard errors in parentheses.

unclear how to compute expected earnings without having an idea of how those expectations are formed. In Section 4 I suggest that they are formed by averaging over network-neighbors' earnings in the host country.

To show, however, that expected wage has an influence on the migration and return decision in the sample and to get an idea of the size of the effect a reliable and simple measure is needed. For this reason I use the difference between real GDP per capita in Mexico and the US and multiply US GDP p.c. by the employment rate.

The coefficient of the expected annual wage difference between Mexico and the US is positive and highly significant for the probability of making a trip. I check whether the marginal effect of the wage difference on the probability to go on a trip differs by whether someone is a potential first-time migrant or has gone on at least one migration before, which is the case. The marginal probabilities from table 2 are used for the behavioral parameter of expected wage in the simulation model.

The effect of the expected wage difference on the return decision should be opposite: The higher the expected wage difference, the lower the probability for return. Indeed, the coefficient for expected wage difference is negative, but only significant in the logit model and not in the probit model (see table 1). Therefore I choose not to include the expected wage difference as a behavioral parameter for the return decision.

Hypothesis 2: Number of previous migrants

Munshi (2003) finds that workers with a network are both less likely to be unemployed and have higher wages. Therefore, migrants tend to go where they know somebody, as shown by Lindstrom and Lauster (2001), Flores-Yeffal and Aysa-Lastra (2011) and Massey and Aysa-Lastra (2011). Previous migrants have an incentive to help the newly arrived to find jobs because this increases the flow of information and trade among migrants, as argued by Stark and Bloom (1985). The help of others decreases assimilation costs for new migrants, as shown for Mexican migrants by Massey and Riosmena (2010). Previous migrants influence potential migrants' decisions through the policy channel as well: immigration policy often includes a family reunification element that permits family members of migrants to immigrate as well. However, Beine *et al.* (2011) estimate the relative importance of the different channels for immigrants to the US in a recent paper and find that the immigration policy channel is much less important than the assimilation cost channel and has decreased in importance since the 1980s.

In sum, the more previous migrants somebody knows the more likely he or she is to migrate. This seems to be true for the sample here as well; the coefficient for the number of family members in the US is highly significant (table 1).

The influence of the number of previous migrants on the migration decision is calibrated in Section 6.

Hypothesis 3: Home preference

Migrants are often assumed to have a preference for consuming home amenities (a home bias like in Faini and Venturini (2008) and Hill (1987)). Everything else held constant, utility is always higher if he or she is at home. The hypothesis is therefore: The stronger someone's home preference, the less likely he or she is to migrate.

Assuming that people are heterogeneous in their home preference, I assign each individual an idiosyncratic home preference parameter. For the home preference parameter I use property ownership in Mexico before first migration as a proxy because people who consider it likely that they will spend their life in the home country are more likely to invest in property there rather than in the host country. Logit and probit regressions of the probability to ever migrate on property ownership, individual controls and community fixed effects before first migration (table 3) show that property ownership before first migration is significantly negatively correlated with becoming a migrant. This confirms the findings by Massey and Espinosa (1997).

I create an index from hectares, properties and businesses owned. The number of hectares owned is transformed to a logscale, then the values from the categories are added. The coefficient of the property index is also negative and significant.

The fact that many, but not all survey respondents seem to have rounded the number of hectares to integers entails problems concerning the continuity of the probability density function of the property index. Therefore values for all respondents are rounded in order to arrive at a discrete distribution to simplify analysis. I use the distribution of the property index value of those individuals that were born between 1955 and 1965 and originate from Central-Western Mexico, because those are the ones used for the simulation.

The obtained distribution of property index values before first migration approximately resembles a negative binomial distribution (mean = .617, variance = .874, overdispersion = .552). However, the null hypothesis that the observed distribution and a negative binomial distribution with the above mean and variance are equal was rejected in a Pearsons chi squared and a log likelihood ratio test.

Therefore the relative frequencies of the property index in the Central-Western Mexico subsample are used as relative frequencies for the home preference parameter h_i. The analysis is confined to values for the property index from 0 to 4, because the proportion of individuals with property index larger than 4 is less than 1%. 57.85% of individuals have a property index of 0, 29.96% have a value of 1, 7.83% have a value of 2, 2.57% of 3 and 1.09% of 4.

The probability to make a migration move in a person-year negatively depends on the prop-

	Prob. to ever be a migrant if before first migration			
	Logit (1) Property categories	**Logit (2) Continuous property index**	**Probit (3) Continuous property index**	**Logit (4) Discrete property index**
Variables	**Coefficients**			
Sex (female =1)	-1.007*** (.028)	-1.015*** (.027)	-.519*** (.014)	-1.001*** (.027)
Family members ever migrated	.316*** (.005)	.315*** (.005)	.178*** (.003)	.319*** (.005)
Hectars owned before first migration	-.003*** (.001)	-	-	-
Pieces of land owned before first migration	-.365*** (.019)	-	-	-
Pieces of property owned before first migration	-.870*** (.012)	-	-	-
Number of businesses owned before first migration	-.528*** (.024)	-	-	
Property index	-	-.599*** (.008)	-.304*** (.004)	-
Property index (reference = 0)	-	-	-	-
Property index = 1	-	-	-	-.876*** (.012)
Property index = 2	-	-	-	-1.390*** (.023)
Property index = 3	-	-	-	-1.67*** (.033)
Property index > 3	-	-	-	-1.620*** (.042)
Constant	-.276*** (.065)	-.326*** (.064)	-.250*** (.036)	-.286*** (.065)
Number of observations	452675	452675	452675	452675
Pseudo R^2	0.221	0.219	0.216	0.220

Table 3: Probability to ever become a migrant if before first migration. County and occupation dummies were used. Robust standard errors in parentheses.

Property index	**Average probability to do trip in person-year**	**z**	$P > \|z\|$
0	.034 (.0004)	79.70	0.00
1	.031 (.0005)	68.06	0.00
2	.029 (.0008)	34.50	0.00
3	.027 (.0011)	23.63	0.00
4	.031 (.0015)	20.12	0.00

Table 4: Average probability to do a trip in a person-year at different levels of the property index. Predictive margins obtained via delta method. Standard errors in parentheses. For the simulation it is assumed that the probability to migrate decreases by .003 for people with home-preference = 1, by .005 for people with homepreference= 2, by .007 if homepreference = 3 and by .003 if homepreference = 4.

erty index, as can be seen in table 1. The average probability to migrate in a person-year was subsequently computed at every level of the property index (see table 4).

Interestingly, property index = 4 increases the probability as compared to property index = 3.

Hypothesis 4: Ties to home

Constant and Zimmermann (2003) find that family reasons are a driving force of repeat migration since migrants incur psychological costs when they are not close to family and friends at home. Those ties to the home country can be understood as relationship capital. It is helpful for the migrant's reintegration into the home community upon return. However, the longer a migrant is away from the home country, the stronger might be the depreciation of home country relationship capital through physical distance. This phenomenon is studied analytically by McCann *et al.* (2010) and found to be empirically relevant for the return decision by de Haas and Fokkema (2011).

This yields the hypothesis: The more family and friends someone has at home, and the stronger the links are with them, the more likely someone is to return.

I consider migrants with at least one trip to the host country and estimate the decrease in likelihood of returning to the home country (for people in the host country) or of migrating again (for people in the home country), taking time since last migration move as explanatory variable. This illustrates the diminishing importance of ties across physical distance over time. A probit regression of the likelihood of making a move in a year on the number of years since the last move yields a negative coefficient that is significant at the 1 % level for both the migration and the return decision (see table 1). The links connecting physically distant network neighbors are therefore assumed to become weaker each period by an amount a. This decrease is not linear but diminishes with the number of years since the last trip; the coefficient of the number of years squared is positive and significantly different from 0 (not shown). The probability of making a move in a person-year (migration or return) starts out at 3.3% when the last trip took place in the previous year. It decreases on average by 1.9% with each additional year that has passed since the last move. After 32 years without a trip the probability is at 1.8%. For reasons of tractability, the effect is assumed to be linear in the simulation. The relationship capital associated with links between physically distant neighbors is therefore assumed to decrease by 2% every year.

The coefficient of the size of the effect of relationship capital in the home country on the probability to return home is estimated in Section 6.1.

Hypothesis 5: Purchasing power

Migrants have a higher purchasing power of their savings in their home country, as modeled by Dustmann (2001). This might be a return motive. Lindstrom (1996) follows a similar argument: He tests whether Mexican migrants from areas which provide dynamic investment opportunities stay longer in the US in order to accumulate more savings that they can put to productive use in their home country and finds some evidence in favor of his hypothesis. Reyes (2004) shows that devaluation of the peso relative to the dollar leads to more return migration, providing another piece of evidence in favor of the purchasing power motive. A related argument is brought forward by Berg (1961) and Hill (1987), who discuss the case where migrants have the objective to achieve a level of lifetime income, and once that is achieved they return home because they have a preference for home country residence. Either argument yields the same conclusion: Holding everything else constant, the higher someone's savings are, the more likely he or she is to return.

Unfortunately, the MMP128 does not provide information on savings. Therefore I try to capture the supposed purchasing power effect by including the last wage in the US, multiplied by the exchange rate of that year and by the consumer price index from the Bank of Mexico, and a dummy that is 1 if property ownership was larger in t+1 than in t in the return regression (see table 1). I also include an interaction term of the dummy and the last wage earnings. The ownership dummy is significant and negative, which implies that people who lived in the US in year t and bought property the same or the following year are less likely to have returned that year than people who did not buy property. That somewhat contradicts the hypothesis and indicates that people seem to buy property rather in the US than in Mexico. The coefficient for the last wage in the US is negative and significant for return, albeit the coefficient is extremely small in size. The interaction term has a positive and significant effect, in line with the hypothesis. This implies that if property ownership in t+1 is larger than in t, the probability of return increases with the wage. The size of the coeffcient, however, is very small as well. For that reason and because the proxy for the purchasing power motive is imperfect I choose not to include it in the model.

Hypothesis 6: Education

Education and heterogeneity in skill levels have been found to be important determinants of self-selection of migrants in a wide range of theoretical papers originating from Borjas (1987) and in empirical studies (Brücker and Trübswetter 2007).

The evidence in the literature on skill selection of Mexican migrants, however, is mixed:

Borjas and Katz (2007), Moraga (2011) and Ibarraran and Lubotsky (2007) find that Mexican migrants to the US are mostly from the lower tail of the Mexican earnings distribution. Other studies find that migrants tend to have a medium position in the country's skill distribution because returns to skill are higher in Mexico, making migration less attractive for highly skilled individuals, while at the same time low-skilled individuals are likely to be more credit-restrained and not able to afford the moving costs (Chiquiar and Hanson 2005, Lacuesta 2006 and Orrenius and Zavodny 2005). There is furthermore evidence that there is a self-selection process happening for migrants within Mexico (Michaelsen and Haisken-DeNew 2011), but not between Mexico and the United States (Boucher *et al.* 2005).

In the simulation model it is difficult to take different levels of education and/or skills into account without significantly increasing the complexity of the problem. The fact that the individuals in the subsample are predominantly low-skilled (81% of migrants born between 1955 and 1965 had completed 9 years of schooling or less) in combination with the very mixed evidence in the literature points in the direction that it does not seem to bias the results dramatically to leave out education and assume a uniform level of education across individuals. This path is chosen here.

Hypothesis 7: Age

All cohorts display a similar migration behavior during their life cycle (see figure 3).

Migration behavior starts around age 18, reaches a peak between the ages of 25 and 30, and then decreases, with small peaks in both migration and return behavior at around age 70. Age might therefore have an effect on migration and return moves, independent of the other motives.

Age is significant in all regressions, except for the fifth age group in the return regression. All in all, the results confirm the inverted u-pattern shown in fig. 3. Considering marginal probabilities, the probability to make a trip increases by .8 percentage points when entering the age-group of 18 to 30, then decreases by 1.1 when entering the age group 31 to 45, and

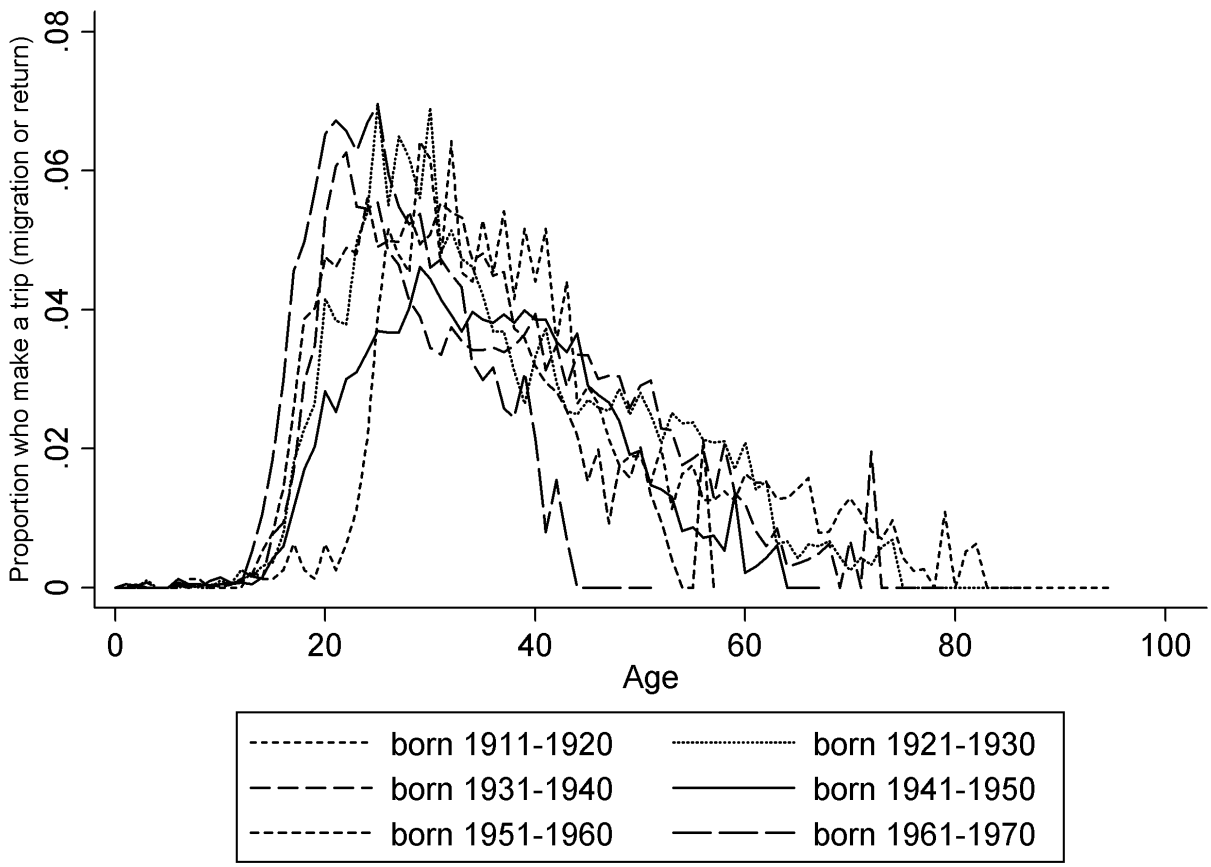

Figure 3: Percent of MMP128 full sample who make a trip at a certain age, for different cohorts.

Age	Marginal probability to do a trip if in Mexico	z	$P > \|z\|$	Marginal probability to return if in the US	z	$P > \|z\|$
< 18	.031 (.001)	26.42	0.00	.304 (.017)	17.53	0.00
18-30	.039 (.001)	73.92	0.00	.367 (.005)	76.26	0.00
31-45	.028 (.0004)	69.70	0.00	.352 (.005)	68.27	0.00
46-60	.022 (.001)	28.57	0.00	.316 (.013)	23.75	0.00
> 60	.009 (.002)	3.96	0.00	.279 (.010)	2.82	0.01

Table 5: Marginal probabilities to make a trip at different ages. Predictive margins obtained via delta method. Standard errors in parentheses.

Parameter	Description	Relevant for	Hypothesis no.	Direction of effect
p_1	Difference expected and current earnings	migration	1	+
p_2	Number of previous migrants in network	migration	2	+
p_3	Home preference	migration	3	-
q_1	Ties to home	return	4	+
p_4, q_2	Age	migration and return	7	mixed

Table 6: Overview of behavioral parameters used.

so on (see table 5). This, being translated to the behavioral parameters in the simulation model, indicates that after 3 periods the agents in the simulation enter the second age group and their probability to migrate increases by .008, and so forth for all age categories up to age 47 (the last year of the simulation) and both migration and return, using the values from table 5.

The behavioral parameters that were included in the model are summarized in table 6.

4 THE MODEL

The model assumes two types of agents, workers and firms, which are spread out randomly on a grid. There are two countries: one with high productivity of labor (the host country), one with low productivity of labor (the home country). Figure 4 displays the initial setup. While workers can move, firms cannot.

Section 4.1 provides an overview of what happens in a single step of the model. Details can be found in Section 4.2.

4.1 Scheduling

The model is initiated via a setup-procedure. During setup, the following happens:

- The world with two countries separated by a wall is initialized.
- Workers are created. A number that is equal to the initial percentage of workers in the home country is assigned a random spot in the home country. The remainder is assigned a random spot in the host country.
- Workers receive their individual values for the home preference and the savings parameter.
- Workers in the home country create links with other workers in their Moore neighborhood, whereas workers in the host country create links with all other workers within a radius s.
- Firms are created in both home and host country and assigned a random spot and a random initial wage.

In every step of a model run the following happens:

- Workers form links to all other workers in their Moore neighborhood (home country) or within a small radius of size s (host country).

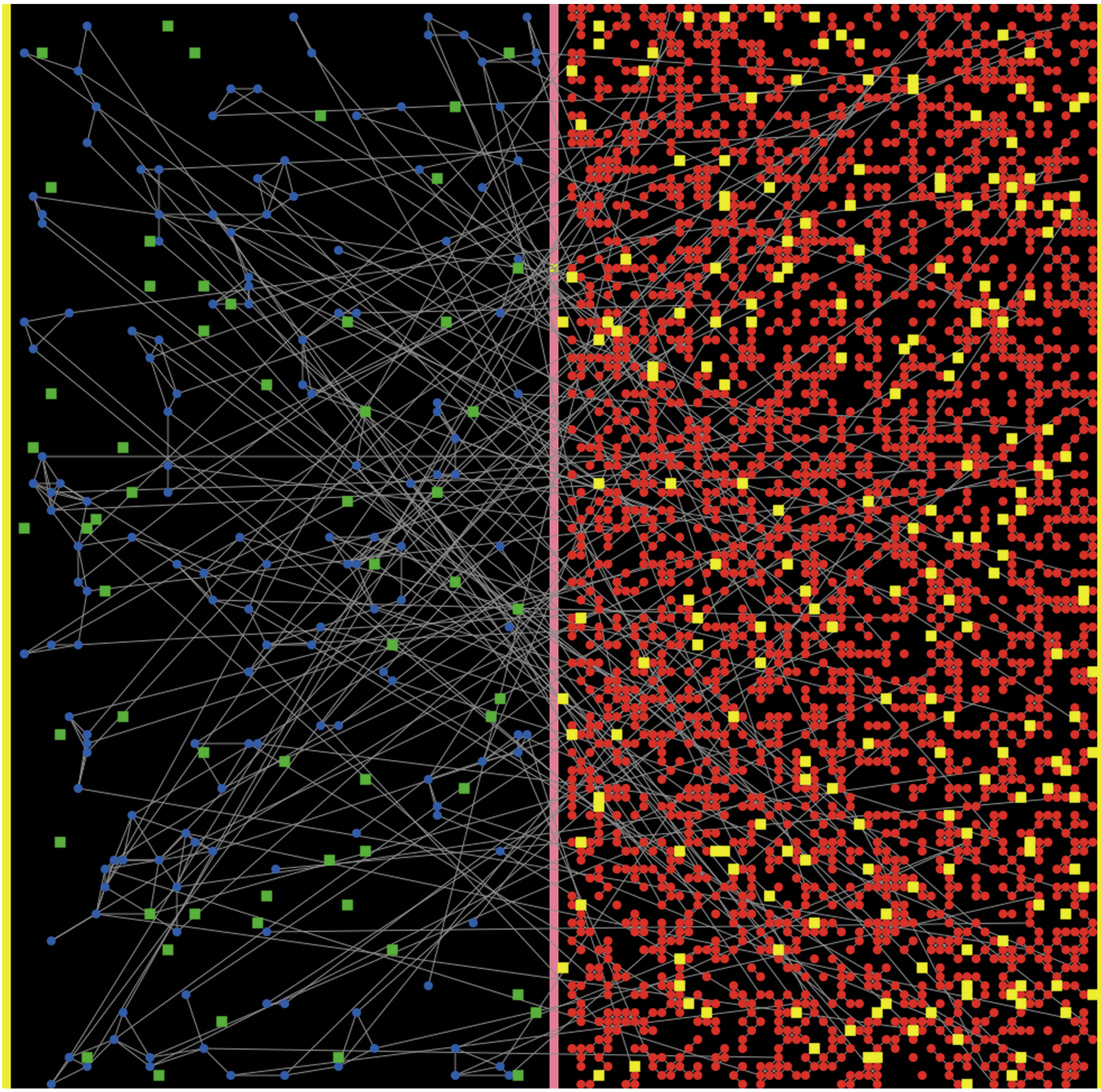

Figure 4: Initial model setup. The model was implemented in NetLogo 4.1.2. (Wilensky, 1999). The complete code is available from the author. The host country is on the left, the home country on the right. Squares are firms, circles are workers, spread out randomly on the grid. Workers are connected by links that represent relationship capital.

- Links between workers that are no immediate neighbors get weaker by amount a (relationship capital diminishing over time due to physical distance).

- All other variables are updated.

- Workers consume their earnings of the previous period minus savings determined by their individual savings rate. Savings are added to wealth.

- Workers without earnings consume a minimum consumption.

- Workers compute their expected wage.

- Migration is a three-step procedure. First, workers in the home country compute whether their wealth is larger than the moving costs. If so, they secondly compute their individual moving probability. Subsequently they draw a random number $\in (0, 1)$. If this number is smaller than their individual probability, they migrate. Their wealth K decreases by the amount of moving costs m_1. In the last step, the probability that somebody who is willing to migrate can do so is determined by the level of border control.

- Migrants become unemployed and decide where to go: If they have any network neighbors in the host country, they move to the network neighbor with the highest wage. If not, they move to a random spot in the host country.

- Unemployed workers in both host and home country move to the network neighbor in the same country who is employed and has the highest wage of all network-neighbors in the country. If they do not have any network neighbors, they move one step in a random direction in search of employment (but never across the border).

- Firms employ unemployed workers that are on their patch. All workers receive the firm's current wage rate.

- Analogous to the potential migrants, potential return migrants in the host country first determine whether their wealth is larger than the return costs and then decide to return with their individual probability.

- Return migrants' wealth decreases by the amount of return costs m_2. They become unemployed and return to the spot in the home country they were assigned in the setup-procedure.

- All measurements of model output take place.

I model the migration history of one generation of heads of household born between 1955 and 1965. If the simulation is started with input data from 1975, when individuals are on average 15 years old, agents start with 0 wealth and first have to accumulate income for one or more periods before they are at least theoretically able to afford the migration costs. This causes them to start migrating much later than their real-world counterparts. Thus, in order to provide the agents with an initial endowment, the simulation is run for 15 periods with the settings of the first period (1975) without allowing the agents to migrate. In the 16th period the schedule as described above starts, output is measured and input data updated in every period. Then, the model is run for 33 time steps, with each step representing one year.

4.2 Model details

Workers form links with all others in their neighborhood, both in the home and the host country. Those links, however, get weaker over time by an amount a per period if both ends of the link are physically separated. Through migration and renewed physical closeness, the relationship capital associated with those links can be replenished, like in McCann *et al.* (2010).

Each period, workers decide whether to migrate (if currently in their home country) or whether to return (if currently in the host country). It is assumed that migrants maximize a utility function that is implicit in the behavioral rules introduced below, rather than explicitly stated. They use heuristics to cope with the high level of uncertainty they face in terms of future earnings, others' migration behavior and future levels of border control.

In every period t, agent i's payoff depends only on the vector of players' actions that particular period, and on the current (payoff-relevant) state of the system (Maskin and Tirole 2001). Workers are heterogeneous only in a home preference parameter (fixed throughout the simulation) and in a savings parameter (time-specific) .

The network enters the migration decision in two ways: First, information is conveyed through it, which has an indirect effect on the probability of migrating. Second, the size of the network increases the probability of migrating directly.

A worker uses the information on earnings of network neighbors in the host country to compute his or her expected earnings in the host country:

$$w_{exp,i,t} = \frac{1}{N} \sum_{n=1}^{N} w_{n,t} \tag{2}$$

where $n = 1, \ldots, N$ are all the worker's network neighbors in the host country, measured at time t.

Workers in the home country only make a migration decision if their wealth K is larger than the moving costs m_1 and if their expected earnings in the host country are larger than their current earnings. The probability that a worker migrates is then a function of the difference between expected and current earnings, the individual home preference parameter, the number of people he or she knows in the host country, and age.

The probability of worker i to migrate at time t is assumed to have the following functional form:

$$\begin{aligned} &p_{i,t}(migrate|K_{i,t} > m_1, w_{exp,i,t} > w_{i,t}) = \\ &p_0 + p_{1,i}(w_{exp,i,t} - w_{i,t}) + p_2 N_i - h_i + p_{4,t} \end{aligned} \tag{3}$$

where $K_{i,t}$ is the worker's wealth in time t, m_1 are the migration costs, p_0 is the baseline migration probability, $p_{1,i,t}$ is the behavioral parameter for the difference between expected and current earnings that depends on whether it is a first migration or not, p_2 is the behavioral parameter for the number of network neighbors in the host country (N_i), h_i is the individual home preference parameter, and $p_{4,t}$ is the age parameter.

In order to keep the model as simple as possible, firms are assumed to pay a fixed, uniform wage to all of their employees. Every step the wage is adjusted exogenously to account for inflation.

The return probability is a function of the number of network neighbors in the home country and the strength of the relationship with them.

The probability of worker i to return at time t given that his or her wealth K is larger than the return costs m_2 is thus assumed to have the following functional form:

$$q_{i,t}(return|K_{i,t} > m_2) = q_0 + \sum_{r=1}^{R} \frac{q_1}{a_{r,t}} + q_{2,t} \tag{4}$$

where q_0 is the baseline return probability, q_1 is the behavioral parameter for ties to the home country, $r = 1, \ldots, R$ are the worker's network neighbors in the home country, $a_{r,t}$ is the age of a link, and $q_{2,t}$ is the age parameter.

5 PARAMETERIZATION OF NON-BEHAVIORAL PARAMETERS

Behavioral parameters that involve the network as well as a baseline migration and return probability, which are not directly measurable, were determined by searching the parameter space for those values that create the closest match between simulated and historical data (see Section 6.1). All other parameters in the model were fixed to empirically determined values (summarized in table 7).

Parameter	**Value used for simulation**	**Source**
City size	6.2	Average county size in California (source: National Association of Counties)
Number people	2,860	Number of Individuals interviewed in the MMP128 survey born between 1955 and 1965 and living in Central-Western Mexico
Initial percentage at home	94.4	Proportion of people from the subset of the MMP that was in Mexico in the year 1975
Border control	annual border enforcement budget normalized to $[0, 1]$	U.S. Immigration and Naturalization Service (until 1998), Homeland Security Digital Library (after 1998), through MMP128 supplementary files
Saving rate	skew-normal distribution with $\xi = 0.616$, $\omega = 0.721$ and $\alpha = -7.5$	ENIGH
Moving costs	1,110.26	MMP128, Instituto Nacional de Estadstica y Geografa
Return costs	1,715.65	US Bureau of Labor Statistics, MMP128
Wage home country	GNI p.c. PPP	World Bank
Wage host country	annual average wage of production and nonsupervisory employees on private nonfarm payrolls	Bureau of Labor Statistics
a: Decrease in relationship capital	2% every period	MMP128

Table 7: Fixed parameters that were derived from data and used for all simulation runs.

I set parameters of the model to sample population parameters that I estimate using the MMP128, the Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) and the Encuesta Nacional de la Dinámica Demográfica (ENADID).

The city size was determined by replicating the ratio of average county size in California (6,969 square kilometers), the most important recipient state, to overall land area of the state (403,933 square kilometers). That yields a radius of 6.2 patches or 49.6 pixels on the grid as radius of a city. This measure determines the radius s in which migrants in the host country build links with other migrants.

The simulation was run with 2,860 agents, the number of heads of household in the MMP128 data set born between the years 1955 and 1965 and interviewed (or having lived in the case of migrants) in the Central-West Mexican states of Sinaloa, Durango, Zacatecas, Nayarit, Jalisco, Aguascalientes, Guanajuato, Colima and Michoacán. Those states together are approximately the size of California and at the same time they comprise the most important states of out-migration. All population distribution measures refer to this subset of the data. The model therefore simulates migration behavior of one generation from one region over the course of 33 years.

The initial percentage of individuals in the home country was set to 94.4 %, the proportion of people from the subset of the data that was in Mexico in the year 1975.

The number of firms in the home country is determined by dividing the number of workers initially in the home country (2700) by the average firm size in Mexico, which according to Laeven and Woodruff (2007) is 13.6 employees per firm. That yields 199 firms. For the host country, the number of firms is assumed to be 58, which is the number of counties in California. In this way, the distribution of migrants across cities can be measured conveniently (see Section 6.2).

Values reported in pesos are converted to USD using the annual average of the official exchange rate for the year the data were measured (reported by the World Bank).

In order to obtain moving costs I compute an average of legal and illegal crossings weighted according to the proportions of legal and illegal crossings in the MMP128 data set. Assuming that real moving costs stay the same over time, I update nominal moving and return costs every period using price indices (the consumer price index by the Bank of Mexico for the moving cost and the consumer price index by the Bureau of Labor Statistics for the return cost). All values reported in the following are at 2002 constant prices.

I assume that (monetary) moving costs for legal migrants are composed of travel costs, costs for obtaining legal entry, and loss of wage income for at least one month around the travel date. This yields a total of 635.79 USD for legal entry (detailed computation available from the author). For illegal entry, the cost for paying the coyote (hired smuggler) is added to the travel cost, which is, considering that some illegal migrants do not use one at all, 801.31 USD at 2002 prices. For simplicity, this cost is assumed to be constant in real terms although in fact it rises slightly with the level of border enforcement (Gathmann 2008). Adding up costs of coyote, travel costs and loss of monthly wage yields a total of 1,314.84 USD for illegal entry. The estimated migration cost is the weighted average of legal and illegal migrations according to the proportions in the subsample of the MMP128 data set (30.13 % legal and 69.87% illegal) and is therefore set at 1,110.26 USD.

Return costs are assumed travel costs plus one month loss of American wages, which are determined by a weighted average of illegal immigrants' and legal workers' wages. The median monthly earnings of a legal worker with Hispanic/Latino ethnicity in 2009 are 1,814.66 USD. I set the estimated earnings of an illegal worker at the average monthly earnings of an illegal migrant in the MMP128 sample, which is 1,548.79 USD. Thus, overall return costs are 1,715.65 USD.

Firms' wages are determined in the following way: In the setup procedure firms are assigned an idiosyncratic productivity parameter $\alpha \sim \mathcal{N}(0, \sigma^2)$ for the host country. The standard deviation $\sigma = .28$ is the standard deviation of the average wage in a county as a percentage of the overall average per capita personal income in California in 2007 from the U.S. Census Bureau. For Mexico, I use the standard deviation of wage across states for the usual Western-

Central Mexican states in 2001 from Chiquiar (2005), which is 22% of overall average wage. Accordingly, each period, a firm's wage is set in the following way:

$$w_{j,t} = \overline{w}_t + \overline{w}_t \alpha_j, \tag{5}$$

where $\alpha \sim \mathcal{N}(0, \sigma^2)$ and $\overline{w}_t$ is the time-specific average wage for the country. For this value I use time-series that are updated each step of the model run. For the US, data from 1975 to 2007 are taken from the average hours and earnings of production and nonsupervisory employees on private nonfarm payrolls by major industry sector data set from the Bureau of Labor Statistics. For Mexico I use GNI p.c. PPP, 1975-2007, from the World Bank, because wage data for the subsample used is not available for all years.

For minimum consumption in the US, I use the average annual expenditure on food and housing of a household in the lowest income quintile of the population in 2010 from the Consumer Expenditure Survey, which is 17,290.81 USD (in 2002 prices). For Mexico, I choose the average annual overall expenditure of a household at the bottom income decile in 2006 from the ENIGH, which is 4,819 USD (2002 prices). For both cases I calculate the percentage of average income (see above) that this value constitutes for the respective year in which it was measured. Assuming that the relation between minimum consumption and average income remains constant over time the minimum consumption is updated by multiplying the average wage each year by .3 for the home country and by .7 for the host country.

To determine the savings parameter, data from the 2008 ENIGH were used. I restrict the data set to 2,860 random observations from Western Mexico, thereby assuming that the sample surveyed for the ENIGH is not in relevant ways different from the one surveyed for the MMP128.

Since only 17 % of respondents make any deposits in saving and other accounts, the difference between current income and current expenditure is used as measure for savings. The distribution of the saving rate in the population is approximately skew normal with parameters $\xi = 0.616$, $\omega = 0.721$ and $\alpha = -7.5$. This distribution is used for the simulation, drawing a savings rate for each worker in each period from this distribution.

Using a principal components analysis I check a set of correlated border enforcement indicators (line watch hours, probability of apprehension, visa accessibility, real border enforcement budget, number of border patrol officers) for principal components in order to find a good proxy for the threshold of border control b, which is the probability of actually being successful when wanting to migrate. Three factors account for 87% of the variance. The border enforcement budget contributes the most to the first factor, which in turn accounts for 54% of the overall variance. The unique variance of the border enforcement budget is one of the lowest as well. Therefore that variable is chosen as a proxy for border control. The annual values from 1975 to 2007 are normalized to $[0, 1]$ so that the probability that an agent who wants to migrate is able to do so is inversely proportional to the level of border enforcement of the respective year. Of course there is a clear endogeneity problem here: if the level of border enforcement is low, a lot of people will decide to try their luck and migrate. That might increase border protection, which in turn influences whether migrants choose to try to cross the border or not. For this reason, the way this is modeled here - migration decision and independent random draw whether migration is permitted - is not realistic. Therefore, I estimate a baseline probability to migrate within the final estimation procedure (Section 6.1) with the border enforcement in place as it is.

6 CALIBRATION AND RESULTS

6.1 Determination of network parameters

The first remaining parameter to be estimated via simulation is the baseline probability to try to make a move in any given year. This cannot be obtained from the data because the data set does not contain information on failed migration attempts of people who end up not migrating at all. I also estimate the baseline return probability via simulation, as well as the two network-related parameters p_2 and q_1. In order to find the best values for the remaining free parameters I run 27,951 combinations of parameters, i.e., I set every parameter to values between 0 and 1 (for p_0, q_0 and p_2) resp. 0 and 2 (for q_1), in steps of .1, and then search around the best values in steps of .01. The parameter combination is determined that is closest to fulfilling three criteria: causing an emergence of the mean, standard deviation and median of the distribution of migrants of cities (see 6.2), causing the emergence of the negative binomial distribution of number of trips of migrants (see 6.3), and yielding a similar time series of flows of migrants and return migrants (see 6.4).

The second stylized fact from Section 2 is an artefact of the model itself: Through the preferential attachment to network neighbors - migrants move to where they know somebody - people from one "hometown" are likely to move to the same host-country location. Accordingly, this criterion is fulfilled for all parameter settings.

The overall best combination turns out to be $p_0 = .1$, $p_2 = .2$, $q_0 = .38$ and $q_1 = .12$ (detailed results of derivation and sensitivity analysis are available from the author).

I subsequently run 10,000 simulations with the best parameter combination found, using different random seeds each time, to see how much the resulting distributions and time series differ from one another and from the empirical ones. All of the following is based on these 10,000 runs with the combination listed above.

6.2 Stylized facts revisited: Migration concentrates on a certain number of places

I directly compare mean, standard deviation and median of the distributions of survey respondents' last US trip and of the last trip of the same number of computer agents and check whether or not the power law hypothesis can be rejected for the simulated data.

To determine the simulated distribution, all patches on the left-hand side of the grid (see fig. 4) with at least one worker on them are brought in a random order. Then, in a radius of city size s, the number of workers is counted that chose this radius as destination for their

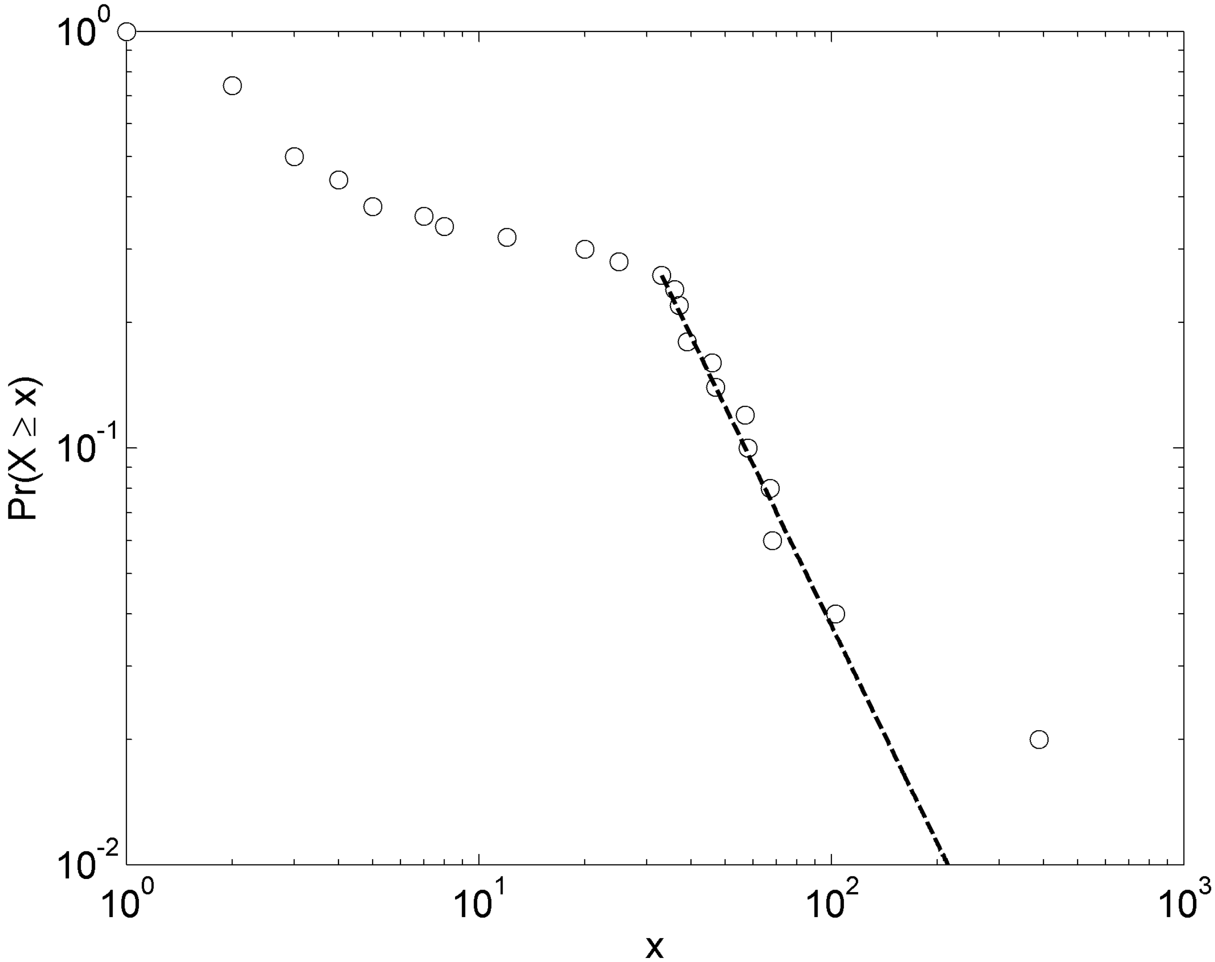

Figure 5: Example of a log-log plot of the cumulative distribution function of numbers of migrants per radius of city size in the simulation with best parameter settings, and fitted values using MLE, with $\gamma = 2.7$ and $x_{min} = 33$. In a Kolmogorov-Smirnov test, $p = 0.68$, so the power-law hypothesis is not rejected.

last migration move. I move on to the next random patch until all workers that migrated at some point have been counted. Finally, the distribution of number of migrants per radius of city size s is determined.

Some of the individual runs are extremely close to the empirically observed mean and standard deviation (e.g. mean = 17.6, standard deviation = 54.9, median = 2 compared to mean = 17.5, standard deviation = 54 and median = 1 for the empirical observation). Just like the empirical distribution of migrants across cities from the small sample, most of the simulated ones also seem to follow a power law (see fig. 5 and fig. 1 for comparison).

However, as for the empirical distributions, one has to be cautious because of the small sample size. The powerlaw hypothesis is rejected for the overall distribution after 10,000 model runs (mean = 27.2, standard deviation = 60.1, median = 4).

Both the facts that the overall simulated distribution after 10,000 runs does not follow a power law and that its median is too high can be ascribed to the fact that there are on average more medium-sized cities in the simulation than in reality. The simulated distribution is not as skewed as the empirical one. This is probably because, for reasons of simplicity, it is not taken into account in the model that some cities attract many more migrants than others not just because of network effects but simply because they are larger and provide better job and other opportunities. Bauer *et al.* (2007) find that the probability that migrants choose a particular US location increases with the total population in that location for almost all groups of migrants. In future versions of this model this should be taken into account.

6.3 Stylized facts revisited: Migration specific capital makes subsequent migration moves more likely

The distribution of the number of trips in the sample was negative binomial. The simulated distribution is not exactly negative binomial because even numbered counts of trips are much more frequent than odd ones in the simulation, but not in reality. That is to say, moving to the host country and moving back at some point is more frequent than in reality. That might be because survey respondents have more degrees of heterogeneity than computer agents: the people who stay in the US are different from the ones who return in a set of characteristics that were not considered here. Furthermore, in reality, some of the migrants have family in the US and others do not, which might fundamentally alter the psychic costs of separation (Lindstrom 1996). Therefore, also their behavioral rules might differ. In the simulation, everyone makes the same type of decision, albeit with different idiosyncratic parameter values such as the home bias $p_{3,i}$. To correct this model inaccuracy in a satisfactory way will be subject of further research. When smoothing the distribution of number of trips by

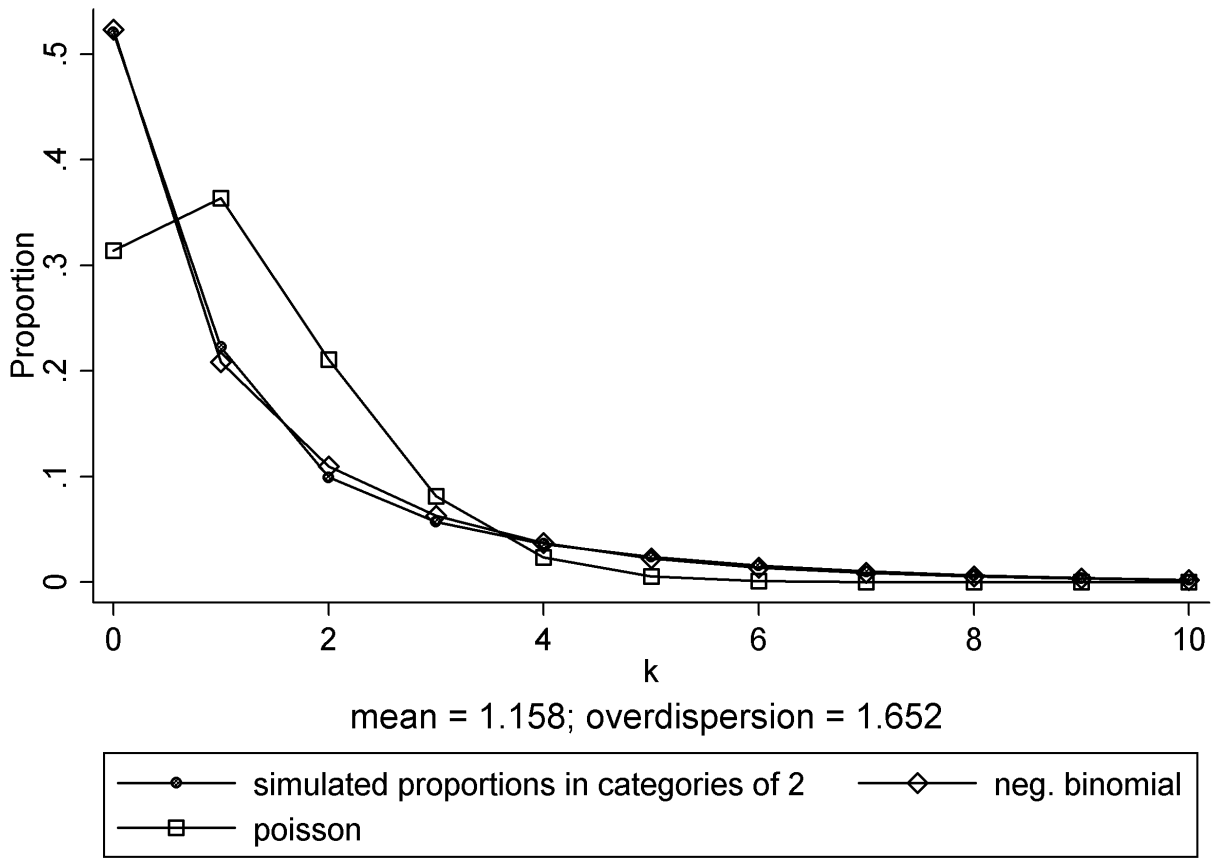

Figure 6: Smoothed distribution of number of trips after 10,000 model runs compared to a Poisson and a negative binomial distribution.

forming categories of two values each to correct for this inaccuracy (0, $1-2$, $3-4$, etc.), the distribution is very close to being negative binomial (see fig. 6).

The hypothesis that the distribution of the smoothed values is negative binomial is, however, rejected in a Kolmogorov-Smirnov test.

6.4 Match of empirical and simulated time series

Generalized least squares was used to find the parameter settings that create the closest match with the observed empirical time series of migration and return (see fig. 7).

In order not to over-calibrate the model the mean squared error between simulated and empirical data was minimized in four points only. Furthermore, I focused on matching

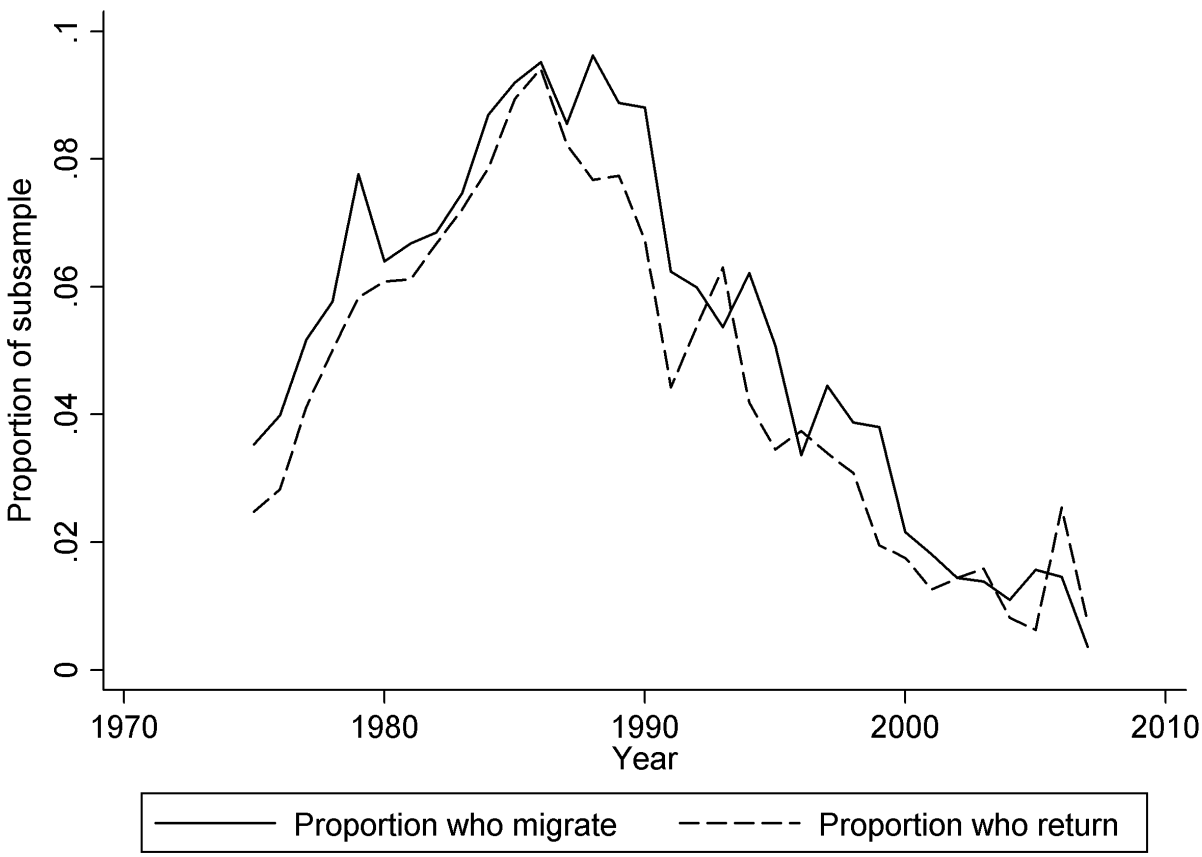

Figure 7: Proportion of MMP128 subsample survey respondents who migrate and return in a given year between 1975 and 2007.

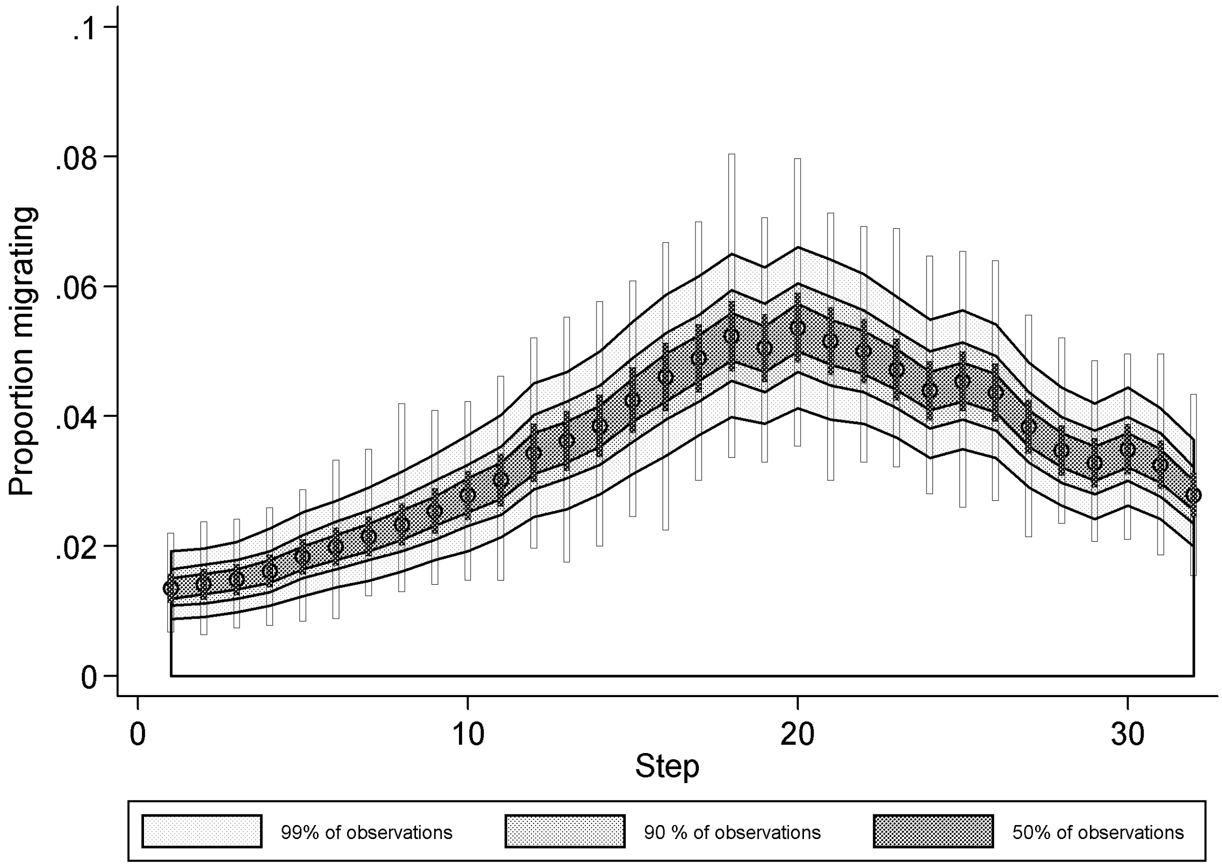

Figure 8: Result of Monte Carlo simulations for proportion of agents migrating. Dark bars: Mean +/- sd; empty bars show range.

rather the overall pattern - an inverted u-shape.

The results of the 10,000 Monte Carlo runs with the best parameter setting $p_0 = .1$, $p_2 = .2$, $q_0 = .38$ and $q_1 = .12$ are depicted in figures 8 and 9. The curves that indicate mean, standard deviation and quantiles can be used to classify particular simulation results in the context of the conceptual population of simulated scenarios, similar to Voudouris *et al.* (2011).

Most of the simulation runs are within a fairly narrow range. The overall pattern - both migration and return movement behavior increase and then decrease over time - follows the pattern of the empirical data.

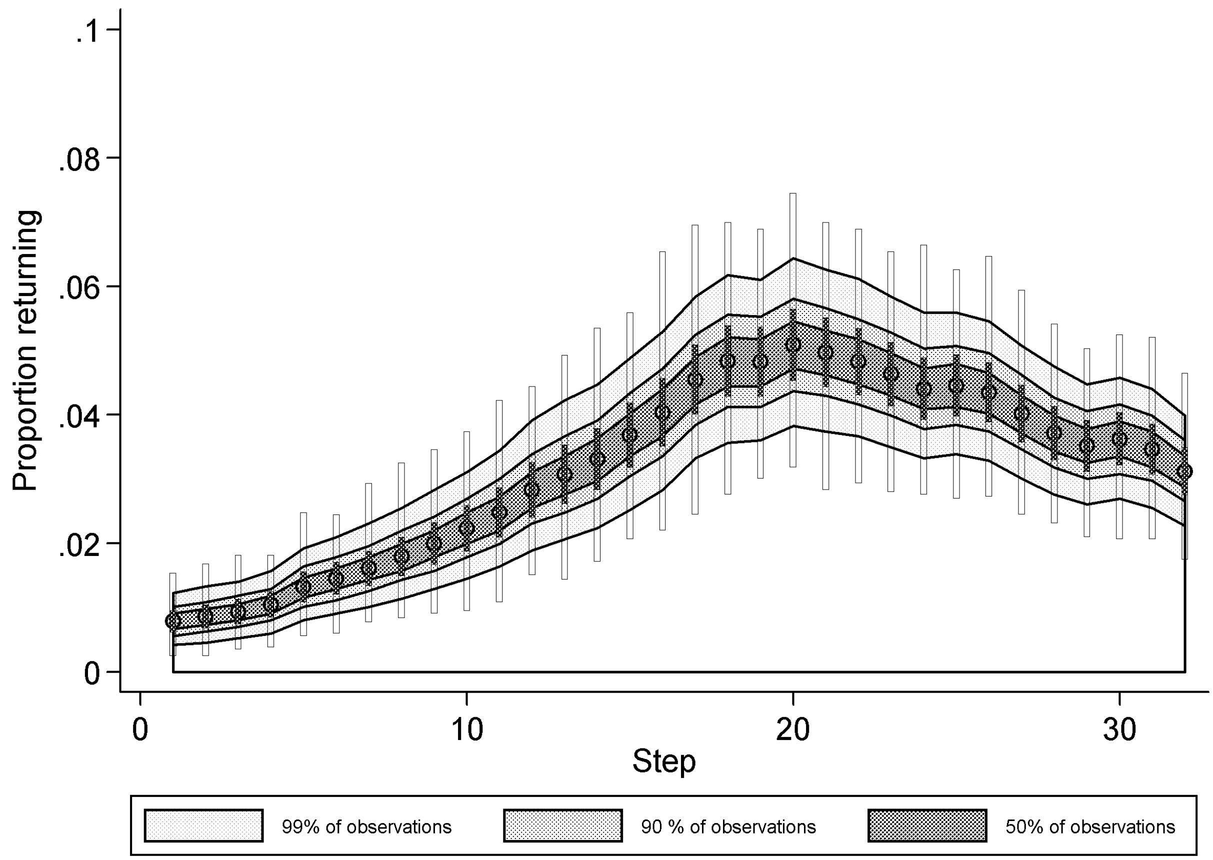

Figure 9: Result of Monte Carlo simulations for proportion of agents returning. Dark bars: Mean +/- sd; empty bars show range.

7 ROBUSTNESS CHECKS AND POLICY ANALYSIS

A robustness analysis of an agent-based model serves to check whether the model reacts as expected when parameter changes are introduced that should alter the results in an unambiguous way. By increasing the home-country wage relative to the host country wage it shall now be demonstrated that the model at hand passes a test for robustness. Afterwards it will be illustrated how the model can be used for policy analysis. It is shown how the effect of a tightening of border control depends on the level of foresight of potential return migrants.

Some potential migrants can probably not afford to migrate and would therefore be enabled to overcome a "poverty trap" if the wages increased slightly (McKenzie and Rapoport 2007). A larger increase in home country wages should decrease stocks of migrants in the host country. To check whether this result is produced by the model, the home country wage is increased by multiplying each value in the time series by $1.1, 1.2...$, up to 3.2 and running the model 1,000 times for each treatment. An increase in average stocks is observed in early periods for increases in the average home wage. At some point every potential migrant has gathered sufficient wealth to overcome the poverty trap. That point is reached the earlier, the higher the home country wage. Beyond that point, the higher the home country wage, the lower are the average stocks in the host country (see fig. 10). At values larger than 3.2, migration ceases almost completely as the home country wages are on average as high as the host country wages.

This is the effect that was expected and it is reproduced by the model. Whether this estimate can be trusted quantitatively depends on whether one believes that the behavioral rules - in particular the impact of the wage difference on the migration decision - are stable if the wages increase substantially. Further research is needed to verify this assumption.

The next experiment that is performed concerns the level of border control. I attempt to

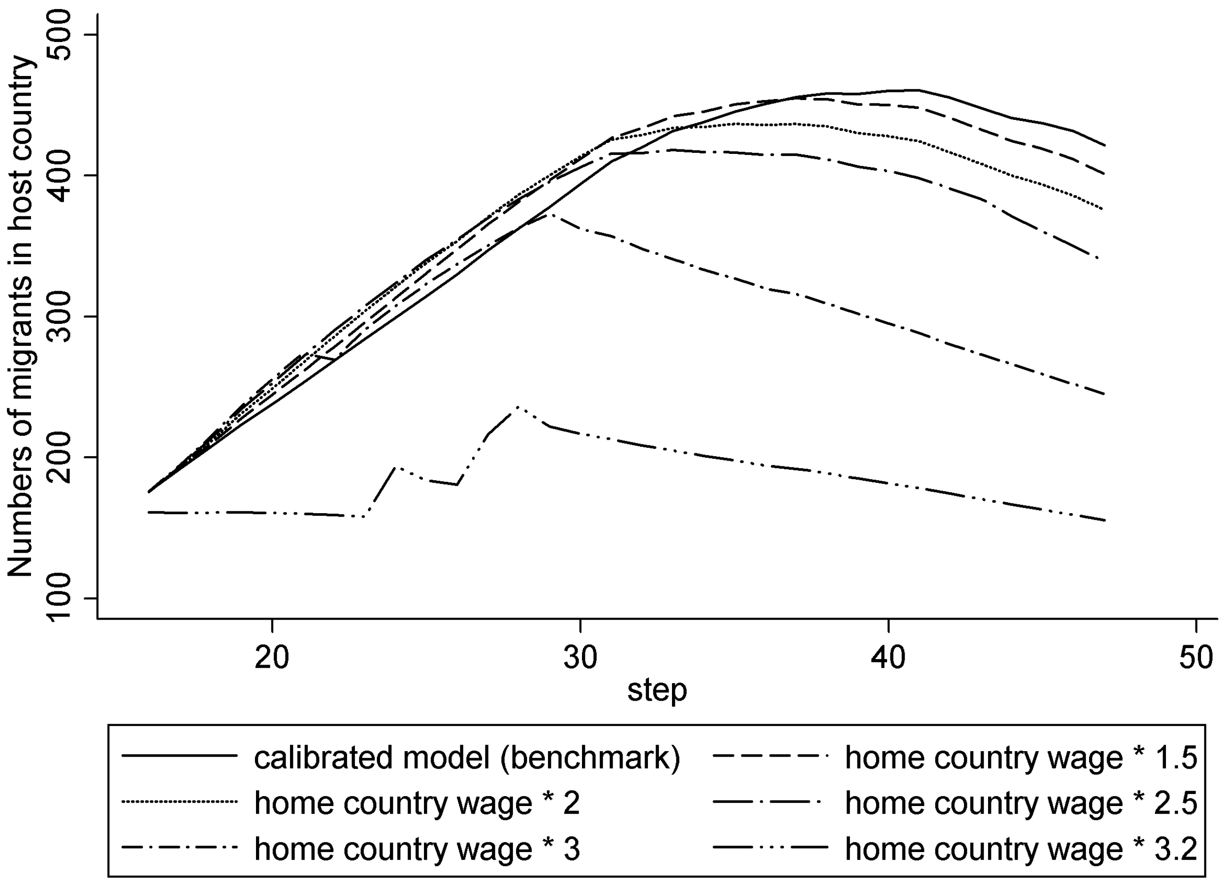

Figure 10: Average stocks of migrants at each model step (1,000 model runs), at different values of the average home country wage.

contribute to the debate whether increasing border protection increases or decreases the stock of migrants in a country. Kossoudji (1992) observes that tighter regulation increases stocks of migrants because it decreases out-migration. Angelucci (2005) finds an ambiguous answer: Tighter border control clearly decreases inflow, but also decreases outflow. Thom (2010) does not find that stricter border control increases stocks of migrants. Clearly, the net effect depends on in how far migrants are deterred from returning since they take into account the lower probability to be able to migrate again.

In order to test the impact on stocks it is assumed that the level of border control increases by 10%. Figure 11 depicts the average stocks across 33 years at levels of baseline return probability of .38 and at lower levels, showing how stocks increase with a decrease in return probability.

The relationship between average stocks and baseline return probability is almost linear. Average stocks increase by 3.58 individuals with every percentage point decrease in baseline return probability. It can be concluded that, on average, stocks increase after an increase in border control by 10% if, of 100 migrants in the US of whom 38 would have returned in a given year, 7 (i.e. 18 %) or more take into account the reduced migration probability and refrain from returning.

8 CONCLUSIONS

In this study I build an agent-based model to analyze the phenomenon of circular migration. It is the first empirically founded and spatially explicit model of the phenomenon that is able to take account of the whole cycle of migration and the role of networks. Three stylized facts on circular migration are introduced that the model can match, despite it being fairly simple: Migration concentrates on a certain number of places, people from one neighborhood tend to go to the same few places, and migration specific capital makes subsequent migration moves more likely. I derive a set of hypotheses from the literature concerning influential

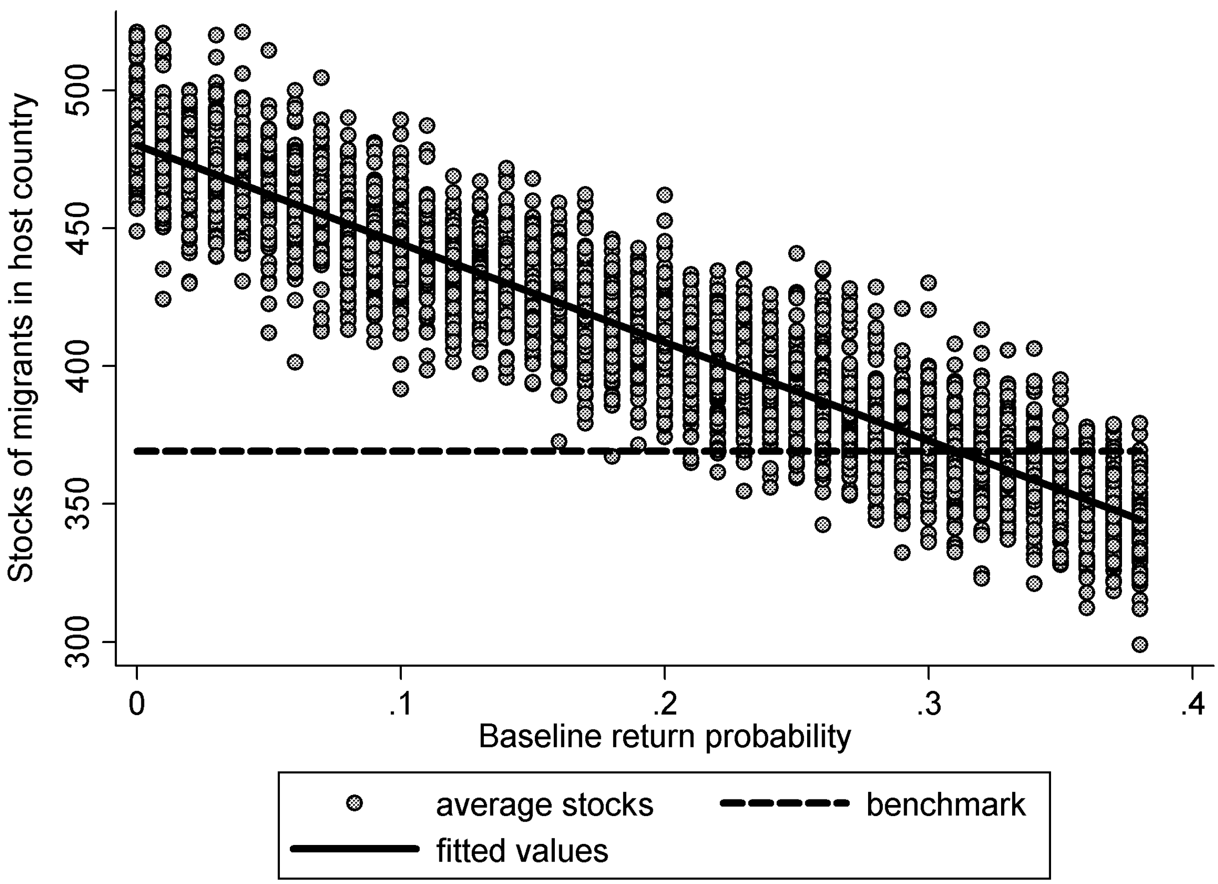

Figure 11: Increase of border control by 10 %: Average stocks of migrants in the host country across 33 years at different levels of baseline return probability. The simulation was run 100 times for each level. The horizontal line indicates the average stocks after 10,000 runs of the benchmark simulation (369.17). The intersection of the fitted values and the benchmark scenario indicates at which level of decrease in baseline return probability the average stocks at the higher level of border control actually start to be higher than in the benchmark scenario.

factors to migrate or to return in a given year. These hypotheses are tested using the Mexican Migration Project (MMP128). The behavioral motives that survived the empirical check are included in the model. I find that expected earnings, an idiosyncratic home bias, network ties to other migrants, strength of links to the home country and age have a highly significant impact on circular migration patterns over time. A model is presented that includes two countries with differing average wages, workers who search for employment, and firms. Workers can migrate and return according to probabilistic behavioral rules estimated from the MMP128. Four remaining parameters are calibrated by running Monte Carlo simulations. Thus, avoiding a common criticism of agent-based models, this model has only 4 degrees of freedom and is yet able to replicate two distributions and two time series from the data fairly well.

Computational experiments are performed in order to demonstrate the robustness of the model. Finally, it is shown how the model can be used to perform policy analysis. It has the potential to help answer the much debated question whether increasing border protection increases or decreases the stock of migrants in a country. I find that if 18 % or more of migrants who would have returned at the lower level of border control take into account that they might not be able to migrate again and therefore refrain from returning, stocks increase. Otherwise, they decrease.

Promising avenues for future research are to make the model spatially accurate using a geographic information system (GIS) or to introduce more sophisticated behavioral rules to account for existing mismatches between data and simulation.

Moreover, with further calibration and sensitivity analysis, the model can be used for forecasting flows of migration and return in certain regions or cities, possibly by combining it with local border enforcement data, and to estimate the effect of labor market shocks or changes in immigration law.

ENDNOTES

Acknowledgement: The author is grateful for helpful comments and suggestions by Simone Alfarano, Lukas Hambach, Wolfgang Luhan, Maren Michaelsen, Matteo Richiardi, Michael Roos, Pietro Terna, Klaus G. Troitzsch, Alessandra Venturini, Vlasios Voudouris, Thomas Weitner, participants at the Eastern Economic Association Conference in New York City in February 2011, the 17th International Conference on Computing in Economics and Finance in San Francisco in June 2011 and seminar participants at Ruhr University Bochum, Philipps-University Marburg and Università di Torino. She would like to acknowledge financial support by the RUB Research School.

REFERENCES

Angelucci, Manuela. "U.S. Border Enforcement and the Net Flow of Mexican Illegal Migration." IZA Discussion Papers 1642, Institute for the Study of Labor (IZA), 2005.

Axtell, Robert. Zipf Distribution of U.S. Firm Sizes. Science 293 (2001): 1818-1820

Bauer, Thomas, Epstein, Gil and Ira N.Gang. The influence of stocks and flows on migrants' location choices. Research in Labor Economics 26 (2007): 199-229.

Beine, Michel, Docquier, Frédéric and Çağlar Özden. "Dissecting network externalities in international migration." CESifo Working Paper 3333, 2011.

Berg, Elliot J.. Backward-Sloping Labor Supply Functions in Dual Economies - The Africa Case. The Quarterly Journal of Economics 75 (1961): 468-492.

Bijwaard, Govert. Immigrant migration dynamics model for The Netherlands. Journal of Population Economics 23 (2010): 1213-1247.

Borjas, George J.. Self-selection and the earnings of immigrants. American Economic Review 77 (1987): 531-553.

Borjas, George J. and Lawrence F. Katz. "The Evolution of the Mexican-Born Workforce in the United States." In Mexican Immigration to the United States, edited by George J. Borjas, 13-55. University of Chicago Press, 2007.

Boucher, Stephen R., Stark, Oded and J. Edward Taylor. "A Gain with a Drain? Evidence from Rural Mexico on the New Economics of the Brain Drain." University of California Working Paper 05-005, UC Davis, 2005.

Brücker, Herbert and Parvati Trübswetter. Do the best go west? An analysis of the self-selection of employed East-West migrants in Germany. Empirica 34 (2007):371-395.

Chiquiar, Daniel. Why Mexico's regional income convergence broke down. Journal of Development Economics 77 (2005): 257-275.

Clauset, Aaron, Rohilla Shalizi, Cosma and M. E. J. Newman. Power-law distributions in empirical data. SIAM Review 51(2007): 661-703.

Constant, Amelie and Klaus F. Zimmermann. "The dynamics of repeat migration: A Markov chain analysis." IZA Discussion Papers 885, Institute for the Study of Labor (IZA), 2003.

Constant, Amelie and Klaus F. Zimmermann. Circular and Repeat Migration: Counts of Exits and Years Away from the Host Country. Population Research and Policy Review 30 (2011): 495-515.

Da Fonseca Feitosa, Flàvia. "Urban segregation as a complex system: an agent-based simulation approach." PhD diss., Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, 2010.

DaVanzo, Julie. Repeat migration, information costs, and location-specific capital. Population & Environment 4 (1981):45-73.

De Haas, Hein and Tineke Fokkema. The effects of integration and transnational ties on international return migration intentions. Demographic Research 25 (2011): 755-782.

Durand, Jorge, Massey, Douglas S. and Rene M. Zenteno. Mexican immigration to the United States: Continuities and changes. Latin American Research Review 36 (2001): 107-127.

Dustmann, Christian. "Why go back? Return motives of migrant workers." In International Migration: Trends, Policy and Economic Impact, edited by Slobodan Djajić, 233-253.

Routledge, New York, NY, 2001.

Entwisle, Barbara, Malanson, George, Rindfuss, Roland R. and Stephen J. Walsh. An agent-based model of household dynamics and land use change. Journal of Land Use Science 3 (2008): 73-93.

Espíndola, Aquino L., Silveira, Jaylson J. and T. J. P. Penna. A Harris-Todaro agent-based model to rural-urban migration. Brazilian Journal of Physics 36 (2006): 603-609.

Faini, Riccardo and Alessandra Venturini. "Development and migration: Lessons from Southern Europe." ChilD Working Papers 10/2008, 2008.

Fernández-Huertas Moraga, Jesús. New evidence on emigrant selection. The Review of Economics and Statistics 93 (2011): 72-96.

Flores-Yeffal, Nadia Y. and Maria Aysa-Lastra. Place of Origin, Types of Ties, and Support Networks in Mexico - U.S. Migration. Rural Sociology 76 (2011): 481-510.

Fontaine, Corentin and Mark Rounsevell. An agent-based approach to model future residential pressure on a regional landscape. Landscape Ecology 24 (2009): 1237-1254.

Gathmann, Christina. Effects of enforcement on illegal markets: Evidence from migrant smuggling along the Southwestern border. Journal of Public Economics 92 (2008): 1926-1941.

Goldstein, Michel, Morris, Steven and Gary Yen. Problems with fitting to the power-law distribution. The European Physical Journal B - Condensed Matter and Complex Systems 41 (2004): 255-258.

Greene, William H.. Econometric Analysis. Upper Saddle River, New Jersey: Prentice Hall, 2003.

Haase, Dagmar, Lautenbach, Sven and Ralf Seppelt. Modeling and simulating residential mobility in a shrinking city using an agent-based approach. Environmental Modelling & Software 25 (2010): 1225-1240.

Harris, John R. and Michael P. Todaro. Migration, unemployment and development: A Two-Sector analysis. American Economic Review 60 (1970): 126-142.

Hill, John K.. Immigrant decisions concerning duration of stay and migratory frequency. Journal of Development Economics 25(1987): 221-234.

Ibarraran, Pablo and Darren Lubotsky. "Mexican Immigration and Self-Selection: New Evidence from the 2000 Mexican Census." In Mexican Immigration to the United States, edited by George J. Borjas, 159-192. University of Chicago Press, 2007.

Kniveton, Dominic, Smith, Christopher and Sharon Wood. Agent-based model simulations of future changes in migration flows for Burkina Faso. Global Environmental Change 21, Supplement 1 (2011): S34-S40.

Kossoudji, Sherrie. Playing Cat and Mouse at the U.S.-Mexican Border. Demography 29 (1992): 159-180.

Lacuesta, Aitor. "Emigration and human capital: who leaves, who comes back and what difference does it make?" Banco de España Working Papers 0620, Banco de España, 2006.

Laeven, Luc and Christopher Woodruff. The quality of the legal system, firm ownership, and firm size. Review of Economics and Statistics 89 (2007): 601-614.

Lessem, Rebecca. "Mexico - U.S. immigration: Effects of wages and border enforcement." New York University, mimeo, 2011.

Lindstrom, David. Economic opportunity in Mexico and return migration from the United States. Demography 33 (1996): 357-374.

Lindstrom, David and Nathanael Lauster. Local Economic Opportunity and the Competing Risks of Internal and U.S. Migration in Zacatecas, Mexico. International Migration Review 35 (2001): 1232-1256.

Makowsky, Michael, Tavares, Jorge, Makany, Tamas and Patrick Meier. "An Agent-based Model of Crisis-Driven Migration." mimeo, 2006.

Manski, Charles F.. Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60 (1993): 531-542.

Maskin, Eric and Jean Tirole. Markov perfect equilibrium: I. Observable actions. Journal of Economic Theory 100 (2001): 191-219.

Massey, Douglas S. and Maria Aysa-Lastra. Social Capital and International Migration from Latin America. International Journal of Population Research 2011 (2011): 18 pages.

Massey, Douglas S. and Kristin E. Espinosa. What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. American Journal of Sociology 102 (1997): 939-999.

Massey, Douglas S. and Fernando Riosmena. Undocumented Migration from Latin America in an Era of Rising U.S. Enforcement. The ANNALS of the American Academy of Political and Social Science 630 (2010): 294-321.

McCann, Philip, Poot, Jacques and Lynda Sanderson. Migration, relationship capital and international travel: theory and evidence. Journal of Economic Geography 10 (2010): 361-387.

Mckenzie, David and Hillel Rapoport. Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. Journal of Development Economics 84 (2007): 1-24.

Mena, Carlos F., Walsh, Stephen J., Frizzelle, Brian G., Xiaozheng, Yao and George P. Malanson. Land use change on household farms in the Ecuadorian Amazon: Design and implementation of an agent-based model. Applied Geography 31 (2011): 210-222.

Michaelsen, Maren and John P. Haisken-DeNew. "Migration Magnet: The Role of Work Experience in Rural-Urban Wage Differentials in Mexico." Ruhr Economic Papers 263, Ruhr-Universität Bochum, 2011.

Munshi, Kaivan. Networks in the modern economy: Mexican migrants in the U.S. labor market. The Quarterly Journal of Economics 118 (2003): 549-599.

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

Radu, Dragos. Social interactions in economic models of migration: A review and appraisal. Journal of Ethnic and Migration Studies 34 (2008): 531-548.

Redner, Sidney. How popular is your paper? An empirical study of the citation distribution. The European Physical Journal B - Condensed Matter and Complex Systems 4 (1998): 131-134.

Reyes, Belinda I.. Immigrant Trip Duration: The Case of Immigrants from Western Mexico. International Migration Review 35 (2001): 1185-1204.

Reyes, Belinda I.. Changes in Trip Duration for Mexican Immigrants to the United States. Population Research and Policy Review 23 (2004): 235-257.

Silveira, Jaylson J., Espíndola, Aquino L. and T.J.P. Penna. Agent-based model to rural - urban migration analysis. Physica A: Statistical Mechanics and its Applications 364 (2006): 445-456.

Stark, Oded and David E. Bloom. The new economics of labor migration. The American Economic Review 75 (1985): 173-178.

Sun, Shipeng and Steven M. Manson. "An Agent-based Model of Housing Search and Intraurban Migration in the Twin Cities of Minnesota." In 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, International Environmental Modelling and Software Society, 2010.

Thom, Kevin. "Repeated Circular Migration: Theory and Evidence from Undocumented Migrants." New York University, mimeo, 2010.

Vadean, Florin and Matloob Piracha. "Circular Migration or Permanent Return: What Determines Different Forms of Migration?" IZA Discussion Papers 4287, Institute for the Study of Labor (IZA), 2009.

Vergalli, Sergio. Entry and exit strategies in migration dynamics. Journal of Labor Research 32 (2011): 362-389.

Voudouris, Vlasios, Stasinopoulos, Dimitrios, Rigby, Robert and Carlo Di Maio. The ACEGES laboratory for energy policy: Exploring the production of crude oil. Energy Policy 39 (2011): 5480-5489.

Wilensky, Uri. "Netlogo." Center for Connected Learning and Computer-Based Modeling.

Northwestern University. Accessed January 11, 2012. http://ccl.northwestern.edu/netlogo/.