Return Migration: an Empirical Investigation

Roman Zakharenko*

December 12, 2008

Abstract

Many people emigrating abroad eventually return home. Yet, little is known about the returnees: who are they and how do they compare to those who did not return? How does their decision to return depend on economic situation at home? In this paper, I empirically analyze the propensity of US immigrants to return. To identify return migration, I use the method adopted from Van Hook *et.al.* (2006). The method is based the U.S. Current Population Survey (CPS) which interviews households for two consecutive years. About a quarter of foreign-born individuals drop out of the sample between the first and the second years, due to various causes including return migration. After eliminating all other causes of dropout, I estimate the propensity of immigrants to return, depending on personal and home country characteristics. I find that the difference between recent immigrants and other immigrants is greater than the difference between men and women, or skilled and unskilled migrants. Thus, assimilation differentiates immigrants more in their decision to return than education or gender. In particular, distance to home country negatively affects return propensity of those who arrived over 10 years ago, and has no effect on recent immigrants.

JEL classifications: O15, F22, J61

Keywords: return migration, panel attrition, assimilation

*International College of Economics and Finance, 11 Pokrovski Bulvar, Moscow 109028 Russia. Phone: 7.495.772.9590, ext. 2296. Email: r.zakharenko@gmail.com

1 Return Migration: an Empirical Investigation

1.1 Introduction

Many emigrants eventually return home. Yet, little is known about the returnees. Are they more or less successful than those who stayed abroad? Does the return propensity increase or decrease with age? Are family ties significant for decisions to return? How do the return patterns depend on their home country culture and economic performance?

In this paper, I analyze empirically the factors affecting return migration. I use Current Population Survey (CPS) data collected by the U.S. Census. This database has three features that make it particularly useful for a study of return migration. First, its size: there are hundreds of thousands person observations available each year. Second, its information on nativity of respondents: the survey identifies immigrants from over 90 world countries and territories, which enables a cross-country analysis. Third, the sample design: each address is questioned several times during two consecutive years, which makes observations longitudinal. By observing respondents prematurely leaving the sample, we can estimate the fraction of immigrants leaving the US, as a function of individual and home country characteristics.

Indeed, an individual may drop out of the sample not only due to emigration, but also due to death, due to moving to another address in the US, and simply due to refusal to continue participation in the survey. All these outcomes cannot be directly identified in the data; in this paper, I develop a methodology for accounting for these causes when the propensity to return home is estimated.

Another problem is that the decision to return is not always voluntary; it is often the case that the US government requires immigrants to leave. This requirement is very likely to depend on personal characteristics, such as education and family ties, as well as the nativity of the immigrant: one might expect that extending their stay in the US is much easier for a person from the UK than for a person from Afghanistan, all other things being equal. Given these considerations, one might think of a supply-demand model: the US government provides the supply of visas or green cards, depending on immigrants' characteristics, and foreign-born individuals demand these visas depending on the same characteristics. Separating supply from demand, however, remains a methodological problem, which was not solved in this paper and left for future research. I only estimate the dependence of observed outcomes on observed immigrant characteristics.

1.1.1 Existing methodology

Overall, return migration is a scarcely studied topic due to lack of data. Data is usually collected within one country, and therefore migrants are not tracked as they move across borders.[1] Given this data limitation, the common approach to approximate return migration is to estimate the number of people which "disappear" from the host country over time. Historically, two methods have been used.

Repeated cross sections With two repeated cross-section nationwide databases (such as decennial US Census), one can use the method developed by Warren and Peck (1980). In economic literature, a version of this method has been used by Borjas and Bratsberg (1996). According to this approach, the entire sample is divided into non-overlapping groups (e.g., immigrants by country of birth). The decrease in the number of migrants within a certain group can be attributed to return migration. Indeed, the researcher should exclude all new migrants, arriving between the two dates (and therefore must observe everyone's year of entry). One observation is thus not an individual, but a subsample of individuals (e.g. all immigrants from Kenya). This method is suitable for studying macroeconomic factors affecting return decisions, but not particularly useful for studying demographic characteristics of return migrants. Indeed, we can disaggregate immigrants by exogenous characteristics (gender, age, year of entry into the host country.[2]) But variable characteristics such as education cannot be controlled for, because individuals can make unobservable transitions from one educational group to another. For example, the number of low-skilled immigrants may

[1]The only known exception is the dataset constructed by German *Institut fur Arbeitsmarks und Berufsforschung* (IAB) which contains information on Turkish migrants returning home from Germany, both before and after their return migration. The study, however, focused only on individuals intending to return, and therefore cannot be used to compare returnees and non-returnees. Since it includes only one home and one host country, it cannot be used for cross-country analysis. Dustmann and Kirchkamp (2002) provide a study based on this dataset

[2]the year of entry is exogenous in the sense that it cannot be changed after the person has immigrated

decrease not only due to emigration or death, but also because some of them have acquired more skill.

One more problem with using repeated Census data is incomplete coverage of the population. In theory, the Census should interview *all* residents of the country. In reality, a small share of population, especially foreign-born population, is not covered. Moreover, the coverage is improving over time, causing a strong downward bias in return migration estimates. For example, the number of people born in country X who entered the US before 1990 must decrease between years 1990 and 2000, due to death and emigration. But due to improved coverage between dates 1990 and 2000, the estimated number of these people may actually increase, resulting in low or even negative return migration estimates.

Panel data With longitudinal/panel data, dropping out of the sample may be attributed to return migration (of course, one has to eliminate other causes of dropping out such as death). The popular sample is German Socio-Economic Panel (GSOEP) which has been used, among others, by Kirdar (2004), Bellemare (2004), Constant and Massey (2003). The strength of GSOEP is that it follows individuals when they migrate within Germany,[3] thus greatly reducing the dropout rate and allowing to identify return migrants more accurately. A shortcoming of GSOEP is that it covers immigrants from relatively few countries (mainly, Southern Europe) and therefore does not allow to study the effect of home country characteristics on the decision to return.

In contrast with Germany, the United States has a large population of immigrants from most world countries, making it the best object of study when cross-(home)country differences in migration patterns are in question. And the largest longitudinal source of data about the US immigrants is the Current Population Survey, making it a natural choice of a researcher interested in return migration patterns: it allows to study the effects of both personal characteristics, and home-country macroeconomic characteristics, on the decision to return. None of other known datasets allows to study the effect of both of these groups simultaneously.

A methodology for estimating return migration using the CPS data was developed in the demographic literature (Van Hook *et.al.* 2006) and, to my knowledge, has not been used in the field of economics. Demographers,

[3] which is not the case in the American CPS

however, are mostly interested in estimating the total number of returnees, as it allows to estimate the population remaining in the US more accurately. In contrast, the main goal of this paper is to estimate not the amounts of return migration, but the factors affecting the decision to return. Hence, the estimation methodology proposed in this paper is considerably different from Van Hook *et.al.* (2006), although the same data source was used.

1.1.2 Existing hypotheses and findings about return migration

Historically, two disjoint sets of hypotheses about return migrants have been discussed: how return migration patterns differ by country of origin, and how they differ by personal characteristics such as age, gender, human capital or job market performance.

The studies of return migration depending on home country characteristics usually find that immigrants are more likely to return to wealthier and to geographically closer countries (e.g. Jasso and Rosenzweig 1982, Borjas and Bratsberg 1996). In this paper, home country GDP is found insignificant for return migration decisions, while distance to home does matter, but not for all groups of immigrants. Borjas and Bratsberg (1996) also find that migrants from Communist countries[4] are far less likely to return than others. This finding hints that large institutional differences between country X and the US makes immigrants from X much less likely to return home from the US. A similar pattern is observed in this paper: immigrants from muslim countries, which have vast institutional and ideological differences from the US, rarely leave the US. At the same time, institutional differences between the US and ex-Soviet countries have diminished, and immigrants from those countries are not much different from other immigrants.

The effect of personal characteristics also received attention in the literature. In this literature, there exist undisputable findings such as: family ties at home increase the likelihood of return; family ties in the host country reduce the likelihood of return; recent immigrants are more likely to return than others. In this paper, these finings are confirmed. At the same time, it is unclear whether more successful or less successful immigrants are more likely to return; Constant and Massey (2003) report a dozen of different studies, with widely ranging results. I find that unskilled migrants return more often; also, personal and home-country characteristics affect skilled and unskilled

[4]they use data on US immigrants in the 1970's

migrants somewhat differently.

The interaction of macroeconomic and personal characteristics, to my knowledge, has never been discussed, due to data limitations. For example, does the difference between male and female immigrants depend on the country of origin? It is quite likely that gender differences for immigrants from OECD countries are not the same as gender differences among those born in muslim countries. Similarly, it may (or may not) be the case that unskilled immigrants from different countries have much higher heterogeneity in return migration decisions than their skilled counterparts. The methodology offered in this paper allows, possibly for the first time, to test such hypotheses.

1.1.3 Return migration vs. emigration to third countries

When estimating the fraction of immigrants leaving the US, we cannot claim that they necessarily return home: part of them could go to third countries. Given limited data availability, it is not possible to estimate accurately how many foreign-born emigrants choose to return home, and how many migrate to third countries. However, a partial inference can be made using Integrated Public Use Microdata Series (IPUMS) which provides large (up to 10% of population) samples collected in several countries of the world. The IPUMS database has a particularly good coverage of the Latin countries: it has data from Mexico, Costa Rica, Colombia, Brazil, Argentina, and Chile, covering most of the Latin world.[5] Information from another likely destinations of Latin foreign-born leaving the US – Spain and Portugal – is also available. Using this information, we can estimate the number of, say, Mexican-born individuals who migrated to the US and then either returned to Mexico (return migrants) or migrated to all other countries listed above (third-country migrants). The same exercise can be done for all other Latin countries listed above. The results are listed in table 1. Mexicans leaving the US rarely travel to third countries; among other countries, about 97% of those leaving the US return home and 3% go to other destinations (Argentina is a notable exception with 90/10 ratio). Given these results, we may conclude that migration to third countries is a rare phenomenon compared to return migration. Throughout the rest of the paper, I ignore migration to third countries and assume that all immigrants leaving the US return home, and

[5]Samples from Ecuador and Venezuela are also available, but they lack data on previous migration experience which is vital for identification of return- and third-country migrants

Table 1: Return migration vs. third-country emigration

Home country	Living in US	returned home from US in past 5 yrs	moved to 3rd country from US in past 5 yrs
Mexico	9.3M	267,000	386
Costa-Rica	76,000	4,820	146
Colombia	526,000	22,370	555
Brazil	223,000	12,600	209
Argentina	131,000	4,310	424
Chile	84,000	5,550	204

Note: possible "third countries" are countries listed in first column (excluding home country), Spain, Portugal.

the terms "emigration of the foreign-born" and "return migration" are used interchangeably.

1.2 The method

1.2.1 The main model

Consider an immigrant i living in the US at date t and making choices that affect his/her status at year $t+1$. Generally, the following outcomes are possible:

- stay in the same house – in this case, the immigrant would be observed twice if a survey visits his/her house in years t and $t+1$;
- move to another address in the US;
- emigrate from the US;
- die – of course, this is not the choice of an immigrant but rather an exogenous random process.

I assume that these outcomes are produced by the following discrete choice model. First, the "nature" chooses whether the individual i dies or lives, depending on his/her personal characteristics and an exogenous random process. The following "mortality" function (by analogy with utility function) is computed:

$$U_{di} = X_i\theta_d + \epsilon_{di}$$

where X_i is a vector of observed personal and home country characteristics, θ_d is a vector of parameters labeled as the "propensity to die", while ϵ_{di} is the unobserved component affecting death incidence. The latter is assumed to be drawn from a known distribution and i.i.d. across individuals.

I assume that the individual dies if $U_{di} > 0$ and lives otherwise. Assuming logistic distribution of ϵ_{di}, the probability that the individual dies is

$$P_{di} = \frac{e^{X_i\theta_d}}{1 + e^{X_i\theta_d}}$$

where θ_d reflects the propensity to die, depending on personal characteristics.[6]

If the individual i lives, he chooses whether to stay in the US or return home. The utility from return is

$$U_{ei} = X_i\theta_e + \epsilon_{ei}$$

where θ_e is the propensity to emigrate, and ϵ_{ei} is unobserved i.i.d. random shock drawn from a normal distribution. The utility from non-return is normalized to zero; thus the individual emigrates if $U_{ei} > 0$ and stays in the US otherwise. Assuming independence of ϵ_{ei} from ϵ_{di},the probability of emigration, conditional on not dying, is

$$P_{ei} = 1 - \Phi(-X_i\theta_e) = \Phi(X_i\theta_e)$$

where Φ is the standard normal cumulative distribution function.

Finally, if i does not emigrate, he/she chooses whether to move to another address in the US or stay in the same residence. The model of moving/not moving choice is analogous to the model of emigration. The propensity to move is labeled θ_m; the unobserved component ϵ_{mi} is normally distributed and independent across individuals. ϵ_{mi} may or may not be independent from

[6]The choice of logistic distribution for ϵ_{di} was motivated by the fact that θ_d was borrowed from another research which used the logit model for estimation, see section 1.2.2 for details.

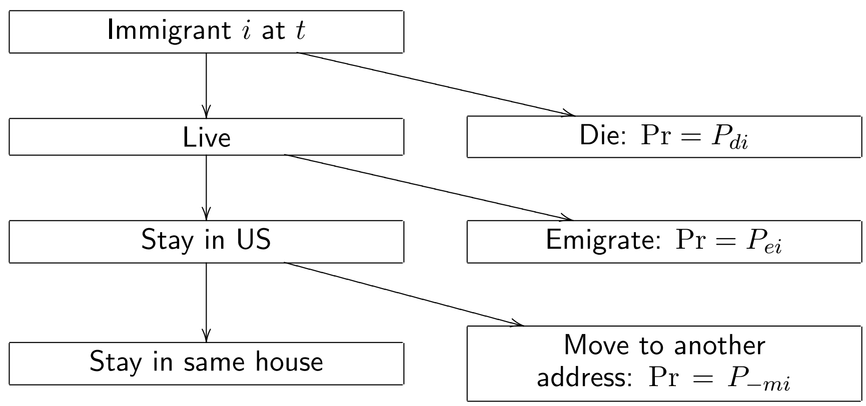

Figure 1: The discrete choice model

$\epsilon_e i$. If independent, the probability of moving, conditional on not emigrating and not dying, is

$$P_{mi} = \Phi(X_i \theta_m)$$

The implications of ϵ_{mi} dependent on $\epsilon_e i$ are discussed in section 1.2.3.

To estimate the model, I use the data provided by the Current Population Survey. Due to the nature of the data, the econometrician only observes whether the person has stayed in the same house one year later (about 73% of all foreign-born) or moved out for whatever reason. In the latter case, the econometrician does not observe what happened to the person: death, or return migration, or moving to another address. Therefore, we have to use some additional sources of information to estimate the chances of death and moving to another address, to figure out the propensity to emigrate θ_e which is out estimation target.

1.2.2 Additional data and model

Death rates The propensity to die, θ_d, was taken from Van Hook *et.al.* (2006) who use data from National Health Interview Surveys and National Health Index to estimate the death rate coefficients shown in table 2.

Mobility within the US To estimate the propensity to move within the US θ_m, we can use the information about recent migration experience of CPS respondents: they report where they lived one year ago. Since the CPS is a representative sample of the US population, the fraction of recent

Table 2: Death rates of immigrants, depending on personal characteristics

Variable	Logit model coefficient
Intercept	-2.017
Male	0.517
Race/Ethnicity (other = 0)	
Mexican	0.336
Other Hispanic	0.067
Non-Hispanic white	0.230
Black	0.176
Age (75 and over = 0)	
18-24	-3.984
25-34	-3.722
35-44	-3.324
45-54	-2.756
55-64	-1.818
65-74	-0.994
Health (poor = 0)	
Excellent	-1.181
Very good	-1.135
Good	-0.937
Fair	-0.586

movers among the CPS respondents approximately equals the true fraction of movers within the US. It is true not only about the *entire* population but also about some groups of population such as foreign-born. Therefore, using recent mobility as a dependent variable in a binary logit model, with personal characteristics X serving as independent variables, should produce a good estimate of θ_m.

One problem related to estimating θ_m arises from the timing of measurements. The model described in section 1.2.1 estimates probabilities of various events within one year, as a function of person characteristics at the *beginning* of the year. The proposed method of θ_m estimation, however, relies on information collected at the *end* of the period. Clearly, some person characteristics might change during the year: age obviously increases by one, educational attainment might improve, citizenship status and employment status might change. All these changes, except age, are not observed and hence cannot be controlled for, creating a possible bias in the estimation of θ_m.

The employment status is the most likely cause of bias because it changes more frequently than educational attainment or marital status or citizenship status, and because it is unclear whether unemployment causes mobility or vice versa. When estimating θ_m, we observe the employment status *after* moving to a new address; the observed positive correlation between mobility and unemployment implies that unemployment may be a *consequence* of recent mobility. However, when estimating the main model described in section 1.2.1, we assume that unemployment is the *cause* of mobility, whether internal or international. Therefore, applying the estimates of θ_m to the main model may produce spurious results. To prevent the problem, I do not use employment status, income, and other volatile characteristics in the regression.

1.2.3 The system of equations: correlated errors

As follows from the above discussion, fitting the model parameters involves estimation of the following system of equations:

$$\begin{aligned} y_i &= I(X_{di}\theta_d + \epsilon_{di} < 0) \times I(X_{ei}\theta_e + \epsilon_{ei} < 0) \times I(X_{mi}\theta_m + \epsilon_{mi} < 0) \quad i \in S_y \\ z_i &= I(X_{mi}\theta_m + \nu_i < 0) \quad i \in S_z \end{aligned}$$

Here y_i is the indicator that the individual i was followed up in year $t+1$, after being first interviewed in year t; I is the indicator function; S_y is the subset of all immigrants of ages 18-70;[7] z_i is the indicator that the person interviewed at t lived in the same house at $t-1$; and $S_z \subset S_y$ includes immigrants of ages 18-70 who lived in the US, either in the same house or at different address, in year $t-1$.

It is generally possible that errors ϵ_{di}, ϵ_{ei}, ϵ_{mi} and ν_i are correlated between each other. The correlation between death error ϵ_{di} and other errors is the least likely and not discussed in this paper; the remaining three errors are more likely to be mutually correlated.

Since the dependent variable in the first equation, y_i, does not depend on ν_i, while the second dependent variable z_i does not depend on $\{\epsilon_{ei}, \epsilon_{mi}\}$, accounting for possible correlation between ν_i and $\{\epsilon_{ei}, \epsilon_{mi}\}$ is not necessary for obtaining consistent estimates of model parameters. But possible correlation between ϵ_{ei} and ϵ_{mi} cannot be ignored, as it may bias the estimate of θ_e.

[7]younger immigrants were excluded because they are unlikely to make individual decisions, and older immigrants were excluded because the probability of death, as well as the error in estimating that probability, is too high

Identifying the correlation between ϵ_{ei} and ϵ_{mi}, however, remains a methodological problem: there are four unobserved errors and only two observed dependent variables, hence the errors cannot be identified.

The problem can be partially solved by accounting for correlation between ν_i and $\{\epsilon_{ei}, \epsilon_{mi}\}$. The following error structure can be specified:

$$\begin{aligned} \epsilon_{ei} &= \phi_e \nu_i + \zeta_{ei} \\ \epsilon_{mi} &= \phi_m \nu_i + \zeta_{mi} \end{aligned}$$

where ζ_{ei} and ζ_{mi} are normal i.i.d. residuals. However, even this assumption does not fully identify the model: we still cannot separate ϵ_{ei} from ϵ_{mi} in the data. For identification of the unknowns, I use the data on another group of the CPS respondents: second-generation Americans (children of immigrants). The following assumptions are made about this group:

- The parameter ϕ_m specified above is the same for the foreign-born and the second-generation Americans. In other words, correlation between past and future mobility within the US is the same for these two groups of respondents. Since second-generation Americans are closer (or, at least, not more distant) to immigrants than any other comparison group, their mobility patterns are the closest to that of immigrants. Indeed, they might still be unequal, but there is no information to identify the difference.

- Second-generation Americans never emigrate. Docquier and Marfouk (2004) report that only 0.4% of US-born working-age population live in other OECD countries. Indeed, some US-born also live in countries other than OECD, but their number is probably even smaller, hence the total number of emigrants should be far below 1% of US adult population. Even if second-generation Americans have a somewhat higher propensity to live outside of the US,[8] these numbers are still incompatible with the estimate of 30% of first-generation immigrants eventually leaving the US (Warren and Peck 1980). With this assumption, we have enough data to identify ϕ_m for the second generation Americans.

1.2.4 The estimation algorithm

The following algorithm estimates the propensity to emigrate, given all above considerations.

[8] from table 6, their likelihood of living abroad is twice as high

First, I estimate ϕ_m, the correlation between past and future internal mobility, using the data on second-generation Americans. I estimate their propensity to move internally, θ_m, by regressing the follow-up indicator y_i on observed characteristics X_i and past mobility indicator z_i.[9] The coefficient for this last regressor, past mobility, is a proxy for ϕ_m, correlation between past and future mobility.

Second, I estimate internal migration propensity θ_m of foreign-born by regressing past mobility z_i on personal characteristics X_i. A simple probit model is used.

Third, I compute the probability P_{mi} that a person moves within the US after the first interview, as a function of observed characteristics X_i and past mobility data z_i. For the former, I use coefficients θ_m estimated in the second step, while for the latter I apply ϕ_m from the first step.

Fourth, I compute the propensity to emigrate θ_e, as a function of personal characteristics and past mobility, after adjusting for the probability of respondent's death and moving within the US.

All estimates are made using the Maximum Likelihood method.

1.2.5 Independent variables: age-period-cohort problem

Among factors that might affect return migration probability, the econometrician may be interested in the following:

- the age of the immigrant. Predicted effect on return migration – uncertain;
- immigrant's duration of stay in the host country: expected to have a negative effect on return probability;
- immigrant's age at entry: should have a positive effect, because younger people assimilate more easily.

These three regressors cannot be directly and simultaneously used in the model because of the collinearity: age equals age at entry plus duration of stay. In the sociological literature this problem, named the age-period-cohort problem,[10] has been discussed since early 1970s, and a number of methods

[9]The probability of death was also accounted for

[10]In classical literature on this problem, the three variables are age, year of observation (period) and year of birth (cohort)

have been developed; see Mason and Wolfinger (2002) for a review. The easiest solution is, indeed, simply to exclude one of these variables from the model. In this paper, I exclude the age of the immigrant and keep duration of stay and age at entry. However, where second-generation Americans are involved, I use their age only because the other two characteristics are not applicable for non-immigrants.

1.2.6 Shortcomings of the method

One disadvantage of this method is that the return migration estimate is not guaranteed to be positive for *all* subsamples of the data. Suppose that some group of people drop out of sample with probability 10%. It may happen that the estimated probability of dropping out for reasons other than migration (that is, moving internally or death) is actually higher than 10%, forcing the emigration probability to be negative. In the logit model, negative probability is impossible; in such cases, the estimation algorithm tries to reduce θ_e down to negative infinity (making emigration probability equal to zero). As a result, computational time greatly increases, and estimates become less accurate. To avoid the problem, I set lower bounds on θ_e parameters.

Another problem is assumed independence of residuals ζ_{ei} and ζ_{mi}. It is generally possible that unexplained willingness to emigrate ζ_{ei} is positively correlated with unexplained willingness to move within the US ζ_{mi}. Accounting for this correlation, however, is impossible, because we do not observe whether the person has died or emigrated or moved if he/she was not followed up.

One more problem is assumed independence across observations: we assume that observations i and j are completely independent from each other. It is most likely not the case if two individuals are members of the same household: their propensity to move or emigrate may be correlated. Since the information on family relationships is available in the CPS data, it is theoretically possible to account for inter-person residual correlations. However, doing so considerably complicates the model; solving this problem is a subject of future research.

1.3 Data

This paper uses data which can be divided into two major categories: person-level and country-level data. The former is information about immigrants in the US, the latter is about their home countries. All data covers years 1998-2007.

1.3.1 Person Data: Current Population Survey

The Current Population Survey is a project administered by the US Census Bureau since early 1940s. Its main goal is to collect data on the US labor force characteristics. Currently, the CPS visits about 100,000 (65,000 before 2002) addresses across all of the US every month. Each month, one-eighth of all addresses are replaced by new randomly chosen addresses, thus each address is visited and interviewed exactly eight times.[11] The visiting pattern is as follows: every address is visited four consecutive months, then left out for eight months, and then visited for four more months. In the dataset, the interviews are numbered by the *month in sample* variable. For example, a household could be visited monthly from February to May 2004 (months in sample 1-4), and then again February to May 2005 (months in sample 5-8). The list of questions asked varies from month to month, but generally consistent across years.

In this study, I use the data collected in March of years 1998-2007. The March survey is the most commonly used by economists and demographers, because it contains the most comprehensive list of socioeconomic questions. Since the interviews are conducted for two consecutive years, each address that was visited in March of year t, must have been also visited either in year $t-1$ or in year $t+1$ (but not both). Consider an example given above: an address visited from February to May 2004, and then again February to May 2005. Since we use March samples only, we observe this household twice: March 2004 (when it was visited for the second time, month in sample = 2) and in March 2005 (month in sample = 6). By observing people living at this address at both dates, we can identify those who have left during the year for whatever reason.

To match person records across years, we have to conduct three steps: first, match addresses across years; second, identify whether an address is occupied by the same household; third, match person records for the same household across years.

[11] except a small number of addresses which became non-residential between visits

Table 3: Number of duplicate address ID's

year	unique ID	2 duplicate ID's	3+ duplicate ID's
1998	64,656	0	3
1999	65,327	38	12
2000	64,857	78	9
2001	64,246	102	14
2002	61,283	33,930	3,635
2003	93,390	6,320	276
2004	93,324	5,372	283
2005	98,664	0	0
2006	97,352	0	0
2007	98,015	0	0

Matching addresses Each address is identified by *household identification number*, which is supposed to be unique for a given combination of sample year and month-in-sample. In practice, however, there are many occasions of duplicate ID's before the year 2005 when the identification methodology was improved. The number of duplicate ID's peaked in year 2002, when only about 60% of ID's were unique. To prevent potential erroneous matches, I dropped all addresses with non-unique ID's. Since the ID's were assigned by the CPS staff, most likely they were not correlated with household characteristics, and therefore dropping ambiguous records should not bias the estimation results. After removing ambiguous records, addresses were matched across years; records without a match were dropped.

Matching households To identify whether the same household lives at an address one year later, the CPS dataset contains the *household number*. In theory, the household number is equal to one during the first interview; in subsequent interviews, it remains the same if the address is occupied by the same household, and increments by one otherwise. In practice, the household number sometimes *decreases* over time (about 0.2% of all addresses), which implies it could be recorded with an error. An erroneous household number could result in both erroneous match (two different households are treated as one) and erroneous mismatch (two records one the same household are treated as different households), causing noise in observations. To account for these errors, I conduct additional checks as described below.

Table 4: Person record matching outcomes, respondents of age 18-70
Percentage points in parentheses

number of matching additional characteristics	same household number	other household number
All three	248,236 (92.85)	555 (1.43)
Two	12,614 (4.72)	3,066 (7.92)
One	4,612 (1.73)	8,820 (22.78)
None	1,887 (0.71)	26,281 (67.87)
Total	267,349 (100.00)	38,722 (100.00)

Matching individuals Usually there are several people living in a household; these people are differentiated by the *line number*. The line number is constant over time for the same person. When a person moves out, the line number is left blank in subsequent interviews. When a new person moves in, he/she is assigned a new (unique) line number. However, when the entire household moves out and is replaced by another household, the line number count starts over. Thus, if two different households were erroneously treated as the same household, the line numbers of two different people could match. To estimate the likelihood of a possible mismatch, I check for consistency of other information supplied by individuals at different dates.

Quality of matching To check whether person records were matched correctly, I check the consistency across years of the following three additional characteristics:

- gender
- age: generally should increase by one. Since the interviews were conducted not *exactly* one year apart, age remaining the same and increasing by two were also accepted
- migration status: "place of residence one year ago" reported in the second year. Respondents should report that they lived in the same place at the time of the first-year interview

The results are presented in table 4.

Overall, 267,349 (78.11%) of all person records could be matched across years according to *household number* parameter.[12] Of them, 92.85% have consistent sex, age, and migration status; 4.72% have a mismatch in one of those characteristics; remaining records have a mismatch in two or all three characteristics.

These results could be produced by erroneously recorded personal characteristics. To approximate the probability of an error in a certain characteristic, I calculate the frequency of a mismatch in this characteristic, conditional on all other characteristics matching. For example, there are 464 observations in which gender doesn't match, while household number, age, and migration status do. Similarly, there are 7,945 (4,205) observations in which age (migration status) is the only mismatching characteristic. Table 4 indicates that there are 555 records with a similar mismatch in the household number: while age, gender, and migration status match (meaning that this is most likely the same person), the household number is different.

Apparently, the household number and gender are much higher quality observations – they are ten times less likely to be recorded with an error. Possibly, there are fewer errors because these characteristics are identified by the interviewer, while age and migration status are reported by the respondent. It is quite likely that the respondent does not remember the exact date of moving to the current residence, or misunderstood the question. It is also possible that the respondent has rounded up his/her age. Figure 2 reports the distribution of respondents' age; there are clearly visible spikes at years 25, 30, 35, etc., which implies that a good number of people are rounding up. It is quite likely that people with certain characteristics (e.g. low education, or foreign-born) are more likely to round up age than others. For example, among natives, 1.98% of all respondents have a mismatch in age (while other characteristics match), while among foreign-born individuals, this figure is 4.43% – more than twice as high! Thus, using age as one of the matching criteria may lead to biased results in the analysis of return migration.

Throughout the paper, I match person records using the household number only. To check for robustness of results, I use an alternative matching rule: person records are matched if at least three out of four characteristics (household number, gender, age, migration status) match. See section 1.4.3

[12] In theory, we should check the consistency of the household number for all household members simultaneously. But for computational speed, all persons were treated independently

Figure 2: Distribution of reported age

for the results of robustness checks.

Description of observed year $t+1$ outcomes For a person observed in year t, the following outcomes can be observed in year $t+1$: (1) person observed again (followed up); (2) person absent, the address is occupied by the same household; (3) the address is occupied by a new household; (4) the second interview could not be conducted (e.g., no one was at home, or the respondents refused to continue participation); (5) the address was vacant in year $t+1$. Outcomes (2), (3), and (5) imply that the person is no longer living in that residence (for whatever reason – death, emigration, or moving to a new address). The fourth outcome is the most problematic: it does not give any information whether the person is still living there or not. The frequencies of these outcome are reported in table 5, for three groups of respondents: natives (US-born respondents with US-born parents), second-generation Americans (US-born with at least one foreign-born parent), foreign-born.

For comparison, table 6 summarizes information on self-reported recent mobility experience of respondents.

The first group of respondents, natives, are the least likely to emigrate: only 0.15% of them returned to the US from abroad within the last year, with probably the same fraction of them moving abroad from the US. Therefore,

Table 5: Observed year $t+1$ outcomes, for respondents of age 18-70
Percentage points

Year-2 observed outcome	natives	second generation	foreign-born
(1) Person followed up	78.91	77.67	73.14
(2) Person absent, same household	5.09	6.02	6.48
(3) Address occupied by other household	5.64	5.87	8.44
(4) No second interview	5.47	5.75	5.82
(5) Vacant address	4.89	4.68	6.11
Number of observations	272,294	22,521	47,450

Table 6: Reported migration experience in the past year, for respondents of age 18-70
Percentage points. Based on year-1 interview

Migration status, 1 yr ago	natives	second generation	foreign born
Lived in same house	85.94	86.38	82.21
Other house in US	13.90	13.33	14.98
Abroad	0.15	0.29	2.81

most of non-followup outcomes (table 5) should be attributed to mobility within the US, and non-followup rates, after some adjustments, should match up with self-reported mobility (table 6). Below, I verify whether the two sources of information about mobility of respondents match up.

Among natives, about 21% drop out of sample before year $t + 1$. Of course, people not responding to the second interview (the fourth outcome in table 5) do not necessarily leave their residence. Assuming that non-response is independent from mobility decisions, I drop those who did not respond; among the remaining population, only 16.5% were not followed up. Of them, some people could die rather than move. Assuming that the death rate for individuals of ages 18-70 is about 0.5%, we end up with about 16% of natives moving from one address to another. On the other hand, table 6 indicates that only about 14% of respondents lived at another address one year ago. The discrepancy in different estimates is about 2%. Given available information, this discrepancy cannot be eliminated.

Birthplace of CPS respondents The CPS asks respondents about their place of birth, which allows to identify immigrants, and about their parents' place of birth, allowing to identify second-generation immigrants. Overall, about 100 distinct home countries can be identified with that data.

Before use, several adjustments had to be made to the birthplace data. First, I exclude observations with too vague birthplace categories such as "other Central America" or "other Africa".

Second, I correct information on birth countries which no longer exist. For example, there are people who describe their birthplace as "Czechoslovakia" and those who were born in "Czech Republic". The criteria of choosing between the two options are not clear; the CPS does not provide any instructions regarding this issue. Table 7 reports the number of immigrants from such countries, disaggregated by the year of immigration.

A problem with dissolved countries of birth is that recently collected home country characteristics (e.g., recent GDP per capita) are not applicable to those countries, and therefore cannot be used as regressors. To handle this problem, I merge immigrants from dissolved countries with immigrants from the most likely successor countries. People born in Czechoslovakia were attributed to Czech Republic, those from Soviet Union were attributed to Russia. The resulting bias is expected to be small, because the number of immigrants with ambiguous birthplace is relatively small, and because

Table 7: Immigrant count, ex-USSR and ex-Czechoslovakia

reported country of birth	immigrated in 1991 or before	immigrated after 1991
USSR	51	87
Latvia	22	9
Lithuania	24	36
Armenia	61	61
Russia	318	484
Ukraine	120	227
Czechoslovakia	68	8
Czech Republic	27	16
Slovakia	22	23

successor countries (Czech Rep. and Slovakia, Russia and Ukraine) have similar institutions and similar economic performance.

1.3.2 Home country data

To study the effect of home country characteristics on return migration, I use several sources of country-level data. Most characteristics are disaggregated not only by country but also by year: this allows to study the effects of not only *levels*, but also *changes* in home country characteristics. Also, with only about 100 home countries observed, one cannot include more than 8-10 country characteristics (some of which are highly correlated) because of regressor collinearity problem. Disaggregation of country data by year greatly increases the number of macro-level observations, allowing to include all observed characteristics into the regression.

The distance between the US and the home country was calculated using the EuGene software (Bennett Stam 2000); it is measured as the shortest distance between national capitals. The missing data was filled manually using Google Earth software.

The economic data was taken from World Economic Outlook database compiled by the International Monetary Fund (IMF data henceforth). Years 1998-2007 were used. The two statistics used were GDP per capita (based on Purchasing Power Parity (PPP), in current prices), and the "exchange rate" defined as the ratio of PPP-based GDP over nominal GDP.[13] This variable was created to verify the hypothesis that the decision to return may

[13] an alternative definition is nominal over PPP exchange rate

be related to a higher purchasing power of the US dollar at home.

Some countries and territories present in the CPS sample are missing in the IMF data (Bermuda, Cuba, Iraq, Puerto-Rico). Information on these countries was taken from PennWorld tables (PWT, Heston Summers Aten 2006). Since this data is available only until 2004, I extrapolated GDP per capita using information on GDP and population growth for these countries; to fill missing exchange rates, I simply extrapolated the last available observation.

Statistics on one more country – Myanmar – was taken from the CIA World Factbook, because neither IMF data nor PWT had reliable information on this country.

The data on the quality of institutions was taken from "Governance" dataset compiled by Kaufmann, Kraay, and Mastruzzi (2006), which provides the following measures: "Voice and Accountability", "Political Stability", "Government Effectiveness", "Regulatory Quality", "Rule of Law", and "Control of Corruption". The data is measured biannually, with the last observation in 2004. The observations in 1999, 2001, and 2003 were imputed by interpolating the 1998, 2000,2002,and 2004 data; the 2005-7 observations are assumed to be equal to that of 2004.

In the Governance database, each country-year observation is made by interviewing a small number of experts that are familiar with the country. The observations are thus made with errors which are estimated by the dataset designers. Typically, the errors are higher in smaller and more remote countries, because fewer experts on these countries could be found. In theory, one has to account for these errors in regression analysis by giving a smaller weight to observations with higher error. For the purpose of this research, however, these measurement errors were ignored: in the CPS data, there are usually fewer immigrants observed from smaller countries; therefore these smaller countries will receive a lower weight in the regression anyway.

Another problem in the Governance data is a very high correlation between some measures. For example, the "rule of law" and "control of corruption" measures have a correlation of almost 97%. Therefore, these measures cannot be used all at once. Throughout the paper, I use the sum of all six characteristics as a measure of institutions.

Besides economic and political measures mentioned above, I include the following country dummies: English-speaking country (a measure of cultural similarity), an OECD country, a transition economy, a muslim country, and a small island country. The list of English-speaking countries was taken from EuGene database, all others were borrowed from Docquier and Marfouk (2005) data. Members of each group are listed in appendix A.

1.4 Results

1.4.1 Benchmark model

The results of the model estimation are given in table 8. Both emigration propensity θ_e and propensity to move internally θ_m are reported.

The propensity to move within the US θ_m has generally predictable patterns. Women, married people, those living in own house are less mobile. Higher education increases mobility within the US. Immigrants become less mobile as they spend more time in the US. The effect of age (age = age at entry + years in the US) is highly negatively significant: the probability that a person moves decreases by about 4% with each extra year of age.

Our main estimation target, the propensity to emigrate θ_e, is presented in the first column of table 8. Since this parameter was basically computed as residual non-followup, many coefficients have a considerably higher standard error than those of θ_m; nevertheless, most of them are significant.

The dependence of θ_e on personal characteristics has a pattern similar to that of θ_m. Women and married people return less often; recent immigrants and those without citizenship status are more likely to leave the US. Higher age at entry increases the likelihood of return, but the effect of current age (-0.029+0.005=-0.024) is negative: older people return less often. It turns out that less educated immigrants are more likely to leave the US. It could be not their own choice but the policy of the US government, which has more restrictive immigration and visa extension rules for those of low skill.

People living in muslim countries have very low rates of emigration to the West, including the United States (see Docquier and Marfouk 2005). From table 8, it follows that they are also far less likely to return – muslim immigrants appear to have made a firm choice not to go back, probably because of institutional differences. Borjas and Bratsberg (1996) made a similar finding about immigrants from Communist countries.[14] Home country institutions do matter in making a decision to return. Not surprisingly, modern immigrants from ex-communist countries (transition economy dummy) do show only a modest difference from the rest of the sample – with changed institutions, migration patterns have also changed and became more "normal".

[14]they used data collected in the 1970s, at the peak of the cold war

Table 8: Benchmark model

regressor	emigration	moving within US
notation	θ_e	θ_m
constant	0.059	-0.337
	(1.488)	(0.461)
person characteristics		
female	-0.167***	-0.061***
	(0.041)	(0.016)
married	-0.437***	-0.073***
	(0.042)	(0.017)
higher education	-0.172***	0.089***
	(0.058)	(0.018)
own house	-0.315***	-0.446***
	(0.045)	(0.017)
health	0.000	-0.017**
	(0.021)	(0.008)
age at entry	0.005**	-0.017***
	(0.002)	(0.001)
years in USA	-0.029***	-0.024***
	(0.003)	(0.001)
non-citizen	0.257***	-0.016
	(0.067)	(0.020)
home country characteristics		
Mexico	0.358***	-0.031
	(0.077)	(0.027)
English-speaking country	-0.136	0.032
	(0.103)	(0.026)
OECD country	0.063	0.070*
	(0.157)	(0.037)
transition economy	-0.283*	-0.083*
	(0.168)	(0.045)
muslim country	-1.030*	0.028
	(0.569)	(0.039)
small island country	0.080	-0.098***
	(0.098)	(0.035)
log(distance to US)	-0.133**	0.003
	(0.061)	(0.016)
log(GDP per capita)	-0.026	0.205
	(0.572)	(0.182)
exchange rate	0.067	0.016
	(0.075)	(0.022)
institutions	-0.018	0.003
	(0.018)	(0.005)
time trend (base=1998)	-0.020**	-0.025***
	(0.009)	(0.003)
past mobility error	0.010	0.323***
	(0.059)	(0.030)
# of observations	41974	40791

* – significant at 90%, ** – at 95%, *** – at 99% level. Standard errors reported in parentheses

It is well known that residents of small island countries are far more likely to emigrate than others (Docquier and Marfouk 2005): one fifth of population living abroad, mainly in the US and Europe, is not uncommon for these countries. One might expect that such high emigration rates lead to shortages in the labor market at home, and eventually to higher return migration rates. According to table 8, this hypothesis is not confirmed: immigrants from small island countries are not any more likely to return than others. However, immigrants from these countries are still different from others; the difference is shown in table 11 and discussed in section 1.4.2.

Immigrants from geographically closer countries are found to me more mobile than others; this is especially true for those from Mexico. Apparently, people from closer countries are more likely to travel back-and-forth than others. People who travel back and forth, obviously, stay in the US for shorter periods and thus more likely to fall in the "recent immigrant" category. If there are more back-and-forth migrants from Mexico, we might expect that the difference in θ_e between recently immigrated Mexicans and other Mexicans is greater than such difference among non-Mexicans. This hypothesis is verified below.

The negative effect of distance to home on return migration was pointed out by Borjas and Bratsberg (1996). However, their other finding, the positive effect of GDP on return migration, could not be confirmed. The exchange rate (the purchasing power of the US dollar in the home country) is also not significant. The effect of institutions, measured as the sum of all six Governance parameters, is of the "wrong" (negative) sign. The insignificance of institutional measures may be due to the fact that there were no major institutional changes in 1998-2004, when the data was collected, and also due to measurement error of the institutional quality.

Overall, we may conclude that economic and Governance institutional characteristics of home countries cannot be used as powerful predictors of migration patterns. Country group dummies providing information about their geography and culture are more powerful determinants of migration patterns.

1.4.2 Emigration by gender, education, length of stay in US

In demographic literature, it is common to treat males and females (especially immigrants) separately, because they are believed to follow very different

Table 9: Emigration by gender

regressor	female	male	difference
constant	-0.105	0.115	-0.219
	(2.138)	(2.003)	(2.929)
person characteristics			
married	-0.436***	-0.428***	-0.007
	(0.063)	(0.057)	(0.085)
higher education	-0.114	-0.189**	0.075
	(0.084)	(0.078)	(0.114)
own house	-0.290***	-0.339***	0.050
	(0.066)	(0.060)	(0.089)
health	0.058*	-0.052*	0.110***
	(0.031)	(0.028)	(0.042)
age at entry	0.008***	0.003	0.005
	(0.003)	(0.003)	(0.004)
years in USA	-0.018***	-0.036***	0.018***
	(0.004)	(0.004)	(0.006)
non-citizen	0.344***	0.164*	0.179
	(0.100)	(0.086)	(0.132)
home country characteristics			
Mexico	0.281**	0.422***	-0.141
	(0.114)	(0.102)	(0.153)
English-speaking country	-0.074	-0.185	0.111
	(0.140)	(0.145)	(0.201)
OECD country	0.123	-0.008	0.131
	(0.210)	(0.229)	(0.311)
transition economy	-0.197	-0.345	0.148
	(0.243)	(0.223)	(0.330)
muslim country	-3.000	-0.578*	-2.422
	(327.219)	(0.318)	(327.220)
small island country	0.016	0.128	-0.111
	(0.135)	(0.138)	(0.193)
log(distance to US)	-0.141*	-0.116	-0.025
	(0.081)	(0.088)	(0.119)
log(GDP per capita)	-0.128	0.004	-0.132
	(0.831)	(0.760)	(1.126)
exchange rate	-0.020	0.118	-0.138
	(0.115)	(0.097)	(0.150)
institutions	-0.030	-0.010	-0.020
	(0.025)	(0.025)	(0.035)
time trend (base=1998)	-0.028**	-0.016	-0.013
	(0.013)	(0.012)	(0.018)
past mobility error	-0.008	0.040	-0.048
	(0.097)	(0.074)	(0.122)
# of observations	21529	20445	

* – significant at 90%, ** – at 95%, *** – at 99% level

patterns. My research, however, did not find vast differences between genders in their return migration pattern; table 9 reports the results.

We can point out the large difference between muslim men and women: the coefficient for muslim women is -3 (the lower bound for this parameter, see section 1.2.6 for explanation). This basically means that they never return; in such cases, the estimates have large standard errors which prevent us from making judgements about significance of these estimates.

Table 10 reports migration differences by educational level. Immigrants from OECD countries stand out against others. Skilled immigrants from OECD are *more* likely to return than other skilled immigrants, which indicates that OECD-US skilled migration is a brain circulation rather than a brain drain. On the other hand, unskilled immigrants from OECD are *less* likely to return than other unskilled emigrants – possibly because of more favorable attitude of the US immigration authorities.

Also, Mexican skilled emigrants stand out against their non-Mexican counterparts much more than unskilled Mexicans do – this finding indicates that returning Mexicans, as well as migrants returning to OECD countries,[15] are subject to positive selection by skill.

Table 11 reports differences between recent immigrants, who spend ten or less years in the US, and those who arrived over ten years ago. The differences between these two groups are far greater than the differences by skill or by gender. In particular, virtually all characteristics indicating the degree of assimilation and sedentariness, such as marital status, house ownership, and citizenship, affect recent immigrants much more than their non-recent counterparts. For example, receiving the US citizenship within the first ten years since immigration greatly reduces the probability of return migration; for those who arrived over ten years ago and still in the US, citizenship plays a smaller role.

The distance to home country differentiates only those who came to the US more than ten years ago; it has no effect on recent immigrants. This finding suggests that distance to home country does not matter *per se*: the decision to return does not directly depend on the cost of return ticket, or on the flight duration. Distance to home may affect migrants in an indirect way: those from more distant countries maintain fewer contacts with home and meet their relatives less frequently; over time, links to home country vanish,

[15] In this research, Mexico was not included into the list of OECD countries, because it has very different migration patterns

Table 10: Emigration by educational level

regressor	skilled	unskilled	difference
constant	0.066	0.193	-0.127
	(2.921)	(1.787)	(3.424)
female	-0.046	-0.183***	0.137
	(0.102)	(0.044)	(0.111)
married	-0.584***	-0.358***	-0.226*
	(0.107)	(0.044)	(0.116)
own house	-0.404***	-0.287***	-0.116
	(0.117)	(0.047)	(0.126)
health	0.037	0.010	0.027
	(0.057)	(0.022)	(0.061)
age at entry	0.010**	0.000	0.010*
	(0.005)	(0.002)	(0.005)
years in USA	-0.169***	-0.019***	-0.151***
	(0.026)	(0.003)	(0.026)
non-citizen	0.502*	0.187***	0.315
	(0.268)	(0.065)	(0.276)
Mexico	0.673***	0.190**	0.484**
	(0.190)	(0.080)	(0.206)
English-speaking country	-0.075	-0.185	0.110
	(0.184)	(0.126)	(0.223)
OECD country	0.564*	-0.394*	0.958***
	(0.296)	(0.216)	(0.367)
transition economy	-0.312	0.109	-0.421
	(0.245)	(0.176)	(0.302)
muslim country	-1.721	-1.401	-0.319
	(1.119)	(2.329)	(2.584)
small island country	-0.053	0.071	-0.124
	(0.242)	(0.103)	(0.262)
log(distance to US)	-0.138	-0.149*	0.011
	(0.099)	(0.080)	(0.127)
log(GDP per capita)	0.088	0.065	0.024
	(1.129)	(0.676)	(1.317)
exchange rate	0.011	0.014	-0.003
	(0.130)	(0.091)	(0.159)
institutions	-0.081**	0.000	-0.081*
	(0.037)	(0.020)	(0.042)
time trend (base=1998)	-0.032	-0.021**	-0.010
	(0.022)	(0.010)	(0.024)
past mobility error	-0.213	0.065	-0.277
	(0.158)	(0.062)	(0.170)
# of observations	17636	24338	

* – significant at 90%, ** – at 95%, *** – at 99% level

Table 11: Emigration by length of stay

regressor	0-10 yrs in USA	over 10 yrs in USA	
constant	-0.370	0.748	-1.119
	(1.930)	(2.178)	(2.910)
female	-0.213***	-0.108*	-0.104
	(0.055)	(0.056)	(0.079)
married	-0.550***	-0.315***	-0.236***
	(0.057)	(0.058)	(0.081)
higher education	-0.064	-0.425***	0.362***
	(0.073)	(0.098)	(0.122)
own house	-0.474***	-0.247***	-0.227**
	(0.069)	(0.058)	(0.090)
health	0.001	0.027	-0.026
	(0.029)	(0.027)	(0.039)
age at entry	0.003	0.003	-0.000
	(0.002)	(0.003)	(0.004)
non-citizen	0.555***	0.129**	0.426***
	(0.152)	(0.062)	(0.165)
Mexico	0.435***	0.157	0.278*
	(0.106)	(0.097)	(0.143)
English-speaking country	-0.244**	-0.003	-0.241
	(0.121)	(0.157)	(0.199)
OECD country	0.152	-0.359	0.511*
	(0.171)	(0.246)	(0.299)
transition economy	-0.320	0.151	-0.471
	(0.206)	(0.219)	(0.301)
muslim country	-1.032*	-0.396	-0.636
	(0.599)	(0.368)	(0.703)
small island country	0.322**	-0.251*	0.574***
	(0.133)	(0.129)	(0.186)
log(distance to US)	0.058	-0.408***	0.465***
	(0.078)	(0.106)	(0.131)
log(GDP per capita)	-0.571	0.273	-0.844
	(0.748)	(0.800)	(1.095)
exchange rate	-0.054	0.203**	-0.258*
	(0.105)	(0.093)	(0.140)
institutions	0.003	-0.022	0.025
	(0.022)	(0.025)	(0.033)
time trend (base=1998)	-0.014	-0.011	-0.002
	(0.012)	(0.012)	(0.017)
past mobility error	-0.082	0.333***	-0.414***
	(0.076)	(0.076)	(0.107)
# of observations	13593	28381	

* – significant at 90%, ** – at 95%, *** – at 99% level

which reduces the incentive to return. In the first few years, however, the ties to home are still strong regardless of distance to home, which makes the distance insignificant factor for recent immigrants.

Table 11 also indicates that duration of stay in the US affects immigrants from small island economies (labeled “islanders” for short) very differently from other immigrants. Recently arrived islanders return *more* often than other recent immigrants. At the same time, non-recent islanders return *less* often than other non-recent immigrants. This finding means that there two very different groups of islanders: those who come to the US for a short period of time, and those who come for good.

It is likely that some immigrants go to the US for a temporary work, and return after they earn enough. One might expect that higher purchasing power of the US dollar at home (labeled as the exchange rate in the regression), makes potential migrants more willing to do so. One might also expect that earning the desired amount of money takes several years of time, and therefore recent immigrants are less affected by the exchange rate. These considerations are confirmed by table 11: a better exchange rate has an insignificant effect on recent immigrants, while it makes non-recent immigrants return more often.

A major methodological problem related to duration of stay in the US is possible selection bias: those who leave are not the same as those who stay, and thus those immigrants who are still in the US after ten years have different characteristics, both observed and unobserved. The latter may become a source of an estimation bias.

The unobserved characteristics are partly captured by recent mobility experience of a respondent. Table 11 indicates that recent mobility within the US affects recent and non-recent immigrants differently: it has no significant effect on former, and a positive effect on the latter. This finding has a plausible explanation: recent immigrants are all mobile by definition, simply because they changed their country of residence in the past few years. Recent mobility within the US does not reveal any new information about them. At the same time, immigrants who arrived over ten years ago may be very heterogenous in their ability to move. Recent mobility experience reveals they are still on the move, and thus more likely to return.

Table 12: Model with alternative matching of person records

regressor	emigration	moving within US
notation	θ_e	θ_m
constant	0.355	-0.337
	(1.041)	(0.461)
person characteristics		
age	0.000	0.000
	(0.000)	(0.000)
female	-0.130***	-0.061***
	(0.029)	(0.016)
married	-0.314***	-0.073***
	(0.030)	(0.017)
higher education	-0.133***	0.089***
	(0.038)	(0.018)
own house	-0.406***	-0.446***
	(0.031)	(0.017)
health	-0.016	-0.017**
	(0.015)	(0.008)
age at entry	0.001	-0.017***
	(0.001)	(0.001)
years in USA	-0.010***	-0.024***
	(0.002)	(0.001)
non-citizen	0.139***	-0.016
	(0.039)	(0.020)
home country characteristics		
Mexico	0.150***	-0.031
	(0.049)	(0.027)
English-speaking country	-0.145**	0.032
	(0.067)	(0.026)
OECD country	-0.268***	0.070*
	(0.099)	(0.037)
transition economy	-0.358***	-0.083*
	(0.136)	(0.045)
muslim country	-0.429***	0.028
	(0.147)	(0.039)
small island country	-0.107*	-0.098***
	(0.064)	(0.035)
log(distance to US)	-0.232***	0.003
	(0.039)	(0.016)
log(GDP per capita)	0.449	0.205
	(0.406)	(0.182)
exchange rate	0.015	0.016
	(0.054)	(0.022)
institutions	-0.016	0.003
	(0.012)	(0.005)
time trend (base=1998)	-0.031***	-0.025***
	(0.006)	(0.003)
past migration error	0.070*	0.293***
	(0.042)	(0.030)
# of observations	41974	40791

* – significant at 90%, ** – at 95%, *** – at 99% level

1.4.3 Robustness

As mentioned in the data description, matching person records across years is not straightforward and several algorithms can be used. Table 12 reports results of a model with alternative method of matching person records: the two records are considered to be records on the same person, if at least three out of four person characteristics match. Generally, coefficients do not change dramatically compared to the benchmark model described in table 8, but some coefficient signs (e.g., OECD country dummy) are reversed. Thus, a more accurate method of matching person records is needed, which is a subject for future work. It is possible to create a model of CPS data collection with explicitly defined error probabilities. These error probabilities can be estimated using the maximum likelihood method; given these estimates, it would be possible to estimate the probability of a match or mismatch of a given person record. The probability of mismatch in each observation could be subsequently used in the estimation of the main model parameters.

1.5 Discussion and future work

This essay utilizes the American Current Population Survey (CPS) to estimate the return migration patterns of US foreign-born. The key feature of the CPS is that each household is interviewed eight times within two years. By using two of these eight interviews, made exactly one year apart, I infer which of the respondents have departed during this year. After adjusting for the probability of death and migration within the US, I estimate what factors affect immigrants' decision to return.

I find that the heterogeneity across recent and non-recent immigrants is greater than the heterogeneity across men and women, or skilled and unskilled migrants. Thus, assimilation differentiates immigrants more in their decision to return than education or gender. In particular, distance to home country negatively affects return propensity of those who arrived over 10 years ago, and has no effect on recent immigrants. This finding implies that distance has no direct immediate effect on foreign-born, but corrodes links to home country, making immigrants from distant countries less willing to return over time. Also, I find that a higher purchasing power of the US dollar in the home country has a positive effect on the return decision, but only for those who have spent a relatively long time in the US.

To improve the estimation methodology, the following things can be done.

First, since the CPS data is collected with errors, the algorithm of matching person records across years could be improved by creating a model of CPS data collection with explicitly defined error probabilities.

Another way to improve the results is to use not only March surveys, but also data collected in all other months. It would allow to increase the number of observations, since many households are visited in months other than March. It would also allow to use eight records on each household instead of two. The latter increases the quality of matching of person records across time: one could identify recording errors more accurately by comparing data from several consecutive months. This strategy would, however, require a more sophisticated matching methodology.

Although a considerable attention was paid to possible correlation of errors in the model, there may still exist unaccounted correlation which may bias the results. Available data does not allow to test or estimate the degree of such correlation. However, it is possible to create a model with a more sophisticated error structure; by making *ad hoc* assumptions about the correlation of errors, I could test whether the model estimates are robust to changing such assumptions.

Finally, it is also possible to include family-level estimation errors into the model, to account for possible correlation of observation errors within the same family.

A Return Migration: an Empirical Investigation

List of English-speaking countries Australia, Bahamas, Barbados, Belize, Canada, Dominica, Fiji, Ghana, Grenada, Guyana, Hong Kong, India, Ireland, Jamaica, New Zealand, Nigeria, Philippines, Puerto Rico, Singapore, South Africa, Trinidad and Tobago, United Kingdom

List of OECD countries Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, South Korea, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom

List of transition economies Armenia, Czech Republic, Hungary, Latvia, Lithuania, Poland, Romania, Russia, Serbia, Slovakia, Ukraine

List of muslim countries Afghanistan, Bangladesh, Egypt, Guyana, Indonesia, Iran, Iraq, Jordan, Lebanon, Malaysia, Morocco, Nigeria, Pakistan, Saudi Arabia, Syria, Turkey

List of small island economies developing economies Bahamas, Barbados, Belize, Bermuda, Cuba, Dominica, Dominican Republic, Fiji, Grenada, Guyana, Haiti, Jamaica, Singapore, Trinidad and Tobago

Table 13: List of countries

country	# of records	country	# of records
Afghanistan	78	Israel	185
Argentina	166	Italy	589
Armenia	97	Jamaica	618
Australia	83	Japan	626
Austria	81	Jordan	70
Bahamas, The	21	Kenya	53
Bangladesh	113	Korea, South	996
Barbados	68	Laos	218
Belgium	63	Latvia	19
Belize	64	Lebanon	175
Bermuda	19	Lithuania	44
Bolivia	70	Malaysia	61
Brazil	318	Mexico	12886
Burma	41	Morocco	45
Cambodia	178	Netherlands	125
Canada	1330	New Zealand	38
Chile	124	Nicaragua	299
China	1233	Nigeria	155
Colombia	735	Norway	38
Costa Rica	86	Pakistan	269
Cuba	1288	Panama	115
Czech Republic	87	Peru	421
Denmark	43	Philippines	2174
Dominica	32	Poland	591
Dominican Republic	1018	Portugal	440
Ecuador	475	Puerto Rico	1799
Egypt	147	Romania	138
El Salvador	1443	Russia	649
Ethiopia	121	Saudi Arabia	30
Fiji	17	Serbia	187
Finland	20	Singapore	29
France	252	Slovakia	30
Germany	1480	South Africa	104
Ghana	115	Spain	151
Greece	223	Sweden	71
Grenada	37	Switzerland	53
Guatemala	671	Syria	66
Guyana	271	Taiwan	392
Haiti	539	Thailand	239
Honduras	446	Trinidad and Tobago	253
Hong Kong	234	Turkey	133
Hungary	116	Ukraine	212
India	1498	United Kingdom	1004
Indonesia	96	Uruguay	64
Iran	407	Venezuela	165
Iraq	135	Vietnam	1112
Ireland	227		

References

[1] Beine, M. , F Docquier, H Rapoport (2001). Brain drain and economic growth: theory and evidence. Journal of Development Economics, vol. 64, issue 1, page 275.

[2] Bellemare, C. (2004). A Life-Cycle Model of Outmigration and Economics Assimilation of Immigrants in Germany, CentER Discussion Paper, no. 29.

[3] Bennett, D. Scott, and Allan Stam (2000). EUGene: A Conceptual Manual. International Interactions 26: 179-204. Website: http://eugenesoftware.org.

[4] Borjas, G., and B. Bratsberg (1996). Who leaves? The outmigration of the foreign-born. The Review of Economics and Statistics, vol. 78, issue 1, pages 165–176.

[5] Cassarino, J.P (2004). Theorising Return Migration: the Conceptual Approach to Return Migrants Revisited. International Journal on Multicultural Societies, vol. 6, no.2, pp. 253-279. UNESCO.

[6] Commander, S., M. Kangasniemi, and L. Alan Winters (2003). The Brain Drain: Curse or Boon? IZA discussion paper no. 809.

[7] Constant, A., D.S. Massey (2003). Self-selection, earnings, and outmigration: A longitudinal study of immigrants to Germany. Journal of Population Economics, vol. 16, no.4, pp. 631-653

[8] Docquier, F., and A. Marfouk (2004). Measuring the international mobility of skilled workers (1990-2000) - Release 1.0. World Bank Policy Research Working Paper, n. 3382.

[9] Dustmann, C., S. Bentolila, and R. Faini (1996). Return migration: the European experience. Economic Policy (April 1996).

[10] Dustmann, C. (2003). Return migration, wage differentials, and the optimal migration duration. European Economic Review, vol. 47, pages 353–369

[11] Dustmann, C. and O. Kirchkamp (2002). The optimal migration duration and activity choice after re-migration. Journal of Development Economics, vol. 67, pages 351-372.

[12] Gould, D.M. (1994). Immigrant links to the home country: empirical implications for U.S. bilateral trade flows. The Review of Economics and Statistics, vol. 76, issue 2, page 302.

[13] Haque, N.U., S.J. Kim (1995). "Human Capital Flight": Impact of Migration on Income and Growth. IMF Staff Papers, vol. 42, No.3, page 577.

[14] Heston, A., R. Summers and B. Aten, Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, September 2006.

[15] Jasso, G., M.R. Rosenzweig (1982). Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States. Demography, Vol. 19, No. 3, pp. 279-290

[16] Jasso, G., M.R. Rosenzweig (1988). How well do U.S. immigrants do? Vintage effects, emigration selectivity, and occupational mobility. Research in population economics, vol.6: a research annual, edited by T. Paul Schultz. Greenwich, Connecticut, JAI Press, 1988. pp. 229-53.

[17] Kaufmann, D., A. Kraay, and M. Mastruzzi (2006). Governance matters V: Aggregate and Individual Governance Indicators for 19962005. World Bank Policy Research Working Paper 4012.

[18] Kirdar, M. (2004). Labor Market Outcomes, Savings Accumulation and Return Migration: Evidence from Immigrants in Germany, mimeo

[19] Ley, D., and A. Kobayashi (2005). Back to Hong Kong: return migration or transnational sojourn? Global Networks, vol. 5 issue 2 (April 2005)

[20] Madrian, B. and L. J. Lefgren (1999). A Note on Longitudnally Matching Current Population Survey Respondents. NBER Working Paper T0247.

[21] Mason, W.M. and N.H. Wolfinger (2002). Cohort Analysis. In International Encyclopedia of the Social and Behavioral Sciences, Elsevier, pp. 151-228

[22] Mountford, A. (1997). Can a brain drain be good for growth in the source economy? Journal of Development Economics, vol. 53, no. 2, pages 287-303.

[23] Olesen, H. (2002). Migration, Return, and Development: An Institutional Perspective . International Migration, vol. 40, page 125

[24] Rauch, J.E. (1999). Networks versus Markets in International Trade. Journal of International Economics, vol. 48, page 7.

[25] Stark, O. (1991). The Migration of Labor. Blackwell, Oxford.

[26] Stark, O. (1996). On the Microeconomics of Return Migration. In V.N. Balasubramanyam and D. Greenaway, Eds., Trade and Development: Essays in Honor of Jagdish Bhawati, pp.32–41. Basingtoke: Macmillan, 1996.

[27] Van Hook, J., W. Zhang, F.D. Bean, J.S. Passel (2006). Foreign-Born Emigration: a New Approach and Estimates Based on Matched CPS Files. Demography, vol. 43, no. 2; p. 361

[28] Yang, D. (2006). Why Do Migrants Return to Poor Countries? Evidence from Philippine Migrants Responses to Exchange Rate Shocks. Review of Economics and Statistics, vol. 88, no. 4, pages 715-735.

[29] Warren, R., J.M. Peck (1980). Foreign-Born Emigration from the United States: 1960 to 1970. Demography, vol.17, no.1, p.71

[30] Zhao, Y. (2002). Causes and Consequences of Return Migration: Recent Evidence from China. Journal of Comparative Economics, vol. 30, pages 376–394