The Determinants of International Migration Accounting for Self-selection *

Simone Bertoli
European University Institute

Jesús Fernández-Huertas Moraga
IAE-CSIC and IZA

Francesc Ortega
Universitat Pompeu Fabra

May 30, 2009

Abstract

Most empirical studies of international migration flows rely exclusively on macro data. However, the literature on internal migration suggests that not accounting for self-selection may be a serious concern. We bridge these two literatures by analyzing a very interesting episode in international migration for which we are able to gather individual-level data covering all relevant countries, namely the exodus of Ecuadorians to Spain and the US in the aftermath of the economic collapse in 1999. Specifically, we produce selection-corrected predictions of counterfactual individual earnings and use them to estimate a discrete-choice migration model that allows for correlated errors across destinations and a rich structure of migration costs. We have three main findings. First, our estimates confirm the need to relax the independence of irrelevant alternatives assumption. Second, the size of migration flows is significantly affected by earnings differentials between the origin and the destinations. Third, our estimates imply that differences in migration costs across destinations are crucial in explaining the observed pattern of migration.

*This paper is part of the INSIDE 2009 projects. Lidia Brun and Feray Koç provided valuable research assistance. Of course, remaining errors are only ours.

1 Introduction

1.1 Motivation

The search for better living and economic conditions induce people to migrate, so that it would seem uncontroversial to claim that differentials in wage levels across destinations deeply influence the choice of potential migrants to move. This entails that migration decisions are responsive to variations in the existing wage differentials. However, providing a convincing empirical assessment of such a simple theoretical implication is harder than it might appear at first sight. Interestingly enough, the economic literature has followed two distinct and markedly different approaches as far as the income responsiveness of international and internal migration decisions are concerned.

The former strand of literature, the international migration literature (e.g. Grogger and Hanson (2008); Belot and Hatton (2008); Ortega and Peri (2009)), addresses this question using macro data, which require to introduce strong implicit assumptions; specifically, country wide average income figures are used to represent potential migrants' expected incomes at destination, and the estimation assumes that no individual-specific unobserved factors simultaneously influences incomes and the decision to migrate. Conversely, the internal migration literature (e.g. Dahl (2002); Bayer, Keohane, and Timmins (2008); Kennan and Walker (2003)) employs individual level data; as an individual is observed in just one among the possible locations, this approach requires first to estimate counterfactual incomes for all the other locations, and then to use predicted incomes as determinants of the internal migration decision. It is a well-established fact in the internal migration literature that failing to account for unobserved ability represents a critical source of bias for the estimation of discrete migration choice models.

1.2 Contribution

Our contribution bridges these two separate strands of literature, by estimating an international migration model using individual level data coming from different countries and sources. The model, which allows for unobserved individual-specific factors to play a role and introduces a rich structure of migration costs, controls for selection using state-of-the-art techniques from the internal migration literature.

This approach is used to analyze a recent major migration episode, namely the wave of

Ecuadorian migration which was triggered by the late 1990s economic crisis, when approximately 600,000 individuals left from a country with a population of 12 million over a few years time (1999-2005). This migration episode, of great interest per se, also offers the chance to address a key challenge, the one represented by data requirements, as recent Ecuadorian migrants moved towards just two main destinations: the United States and Spain. We merge information on Ecuadorians contained in three comparable household surveys collected in Ecuador, the Encuesta Nacional sobre el Empleo y Desempleo en el Area Urbana y Rural 2005 (ENEMDU 2005), the US, the American Community Survey 2007 (ACS 2007; Ruggles, Sobek, Alexander, Fitch, Goeken, Hall, King, and Ronnander (2008)), and Spain, the Encuesta Nacional de Inmigrantes 2007 (ENI 2007). This gives us a sample of 41,335 stayers and 1,786 migrants to the two destinations during this recent wave of migration.

Notwithstanding the large difference in per capita income levels between the two destinations[1] and the existence of stronger prior migration networks in the US (Jokisch and Pribilsky (2002); Bertoli (2009)), the scale of the flow of Ecuadorian migrants to Spain over this period was three times larger than the one to the US. This entails that the analysis of this migration episode can be regarded as a challenging test for any income-maximization theory of migration, which will need to allow for a flexible enough structure of migration costs to reconcile the observed sorting of migrants across the two main destinations with the existing income differentials. Neither a linear nor a log utility model can, in their simplest forms, explain the basic features of the Ecuadorian migration episode in terms of its scale and of the pattern of migrants' selection.

1.3 Main results

Our econometric analysis confirms the empirical relevance of the argument that unobserved individual specific factors need to be adequately addressed in the choice of the estimation procedure, as the test for the independence of irrelevant alternatives is rejected when estimating migration decisions with a simple conditional logistic model on our joint dataset. Furthermore, the application of the selection-control procedure proposed by Dahl (2002) demonstrates that the non-random selection in unobservables biases the counterfactual incomes that we would obtain from simple Mincerian regressions, and this would consequently bias the estimation of the migration decision model. The estimated coefficients of the prop-

[1]The GDP per capita in 2006 US Dollars was 44,000 in the US, 29,000 in Spain and 7,000 in Ecuador.

erly derived income variable from the trinomial probit that is used to model the discrete migration decision demonstrate that migration decisions respond significantly to income differentials across destinations. This holds true independently on whether income is assumed to enter the utility function in a linear or a logarithmic form, suggesting that both competing specifications are supported when estimating individual level migration decisions. The implicit migration costs that we recover from the model show that the cost of migrating to US is several times larger than the corresponding cost of moving to Spain. This differential could be traced back to the cultural and linguistic proximity between Ecuador and Spain and to the relatively generous Spanish welfare state. However, we provide evidence that suggests that the effects of the progressive tightening of the US immigration policies, which began in the mid 1990s together with the relatively lax Spanish immigration policy towards Ecuadorians (at least until 2003) are the most likely reason. Network effects should in principle make migration to the United States cheaper so that they do not seem like a sufficient explanation.

In any case, the variability of migration costs across gender and educational groups confirms that a deeper understanding of the determinants of migration costs is needed in order to more accurately model migration decisions, as Hanson (2008) argued.

1.4 Literature review

This paper is related to four different and, in some cases, unrelated strands of the literature. First, this is a paper on international migration that contributes to the literature on the determinants of migration at the international level. Recent papers who have dealt with this topic following different approaches are the already mentioned Grogger and Hanson (2008), Belot and Hatton (2008), Ortega and Peri (2009), Mayda (2008) and Clark, Hatton, and Williamson (2007). None of these macro papers could control for selection or estimated counterfactual incomes that potential migrants could earn at their destination or origin. Only Clemens, Montenegro, and Pritchett (2008) address this issue by calculating the wages of identical workers with the same nationalities across countries although they do not go on to estimate a migration decision equation.

This step (estimating migration decision models controlling for selection) has a long tradition in the internal migration literature: Nakosteen and Zimmer (1980), Falaris (1987) and Falaris (1988) constitute earlier contributions in this sense. More recent papers are Dahl

(2002) or Bayer, Keohane, and Timmins (2008).

The third strand of the literature is the one on selection methods, clearly intertwined with the previous one. A seminal paper in this respect was Heckman (1979), followed by Lee (1983) and Dubin and McFadden (1984) among others. More recent papers that contributed to this literature are the already mentioned Dahl (2002), Bourguignon, Fournier, and Gurgand (2007) and Bayer, Khan, and Timmins (2008).

Finally, this paper contributes to the literature trying to understand the Ecuadorian migration episode and the crisis that generated it. Some relevant papers in this area are: Jokisch and Pribilsky (2002), Bertoli (2009), Gratton (2007) or Jácome (2004).

The rest of the paper is structured as follows. Section 2 introduces a simple migration decision model which allows for unobserved individual specific factors across individuals. Section 3 outlines an estimation approach that is consistent with our underlying theoretical model. Section 4 briefly sketches the salient features and the economic determinants of the international migration episode that we focus on, namely the Ecuadorian migration that followed the 1999 crisis. Section 5 describes the data sources that we draw upon to build our joint dataset, and it presents the relevant descriptive statistics on the scale of migration flows and on the observed pattern of selection and sorting. Then, Section 6 deals with the implementation of the estimation methodology, Section 7 discusses the results from our individual-level estimation of the income responsiveness of international migration decisions and Section 8 analyzes the pattern of the implicit migration costs that can be recovered from the estimates of the our model. Finally, Section 8 draws the main conclusions of the paper.

2 The Model

We consider the following version of the Roy (1951) migration model. The individuals in a location ($j = 0$) choose whether to migrate to one of either two potential destinations ($j = 1, 2$). Each individual compares the utility from migrating to each destination to the utility from staying in the country of origin and makes the utility-maximizing choice. Individuals differ in observable and unobservable characteristics. More formally, our empirical model has two inter-related equations: a discrete migration-choice equation (1a, 1b) and a wage equation (2). That is, for each location $j = 0, 1, 2$,

$$U_{ij} = \alpha w_{ij} + x_i'\beta_j + (\lambda_j\sigma_i + \epsilon_{ij}^m) = \alpha w_{ij} + x_i'\beta_j + v_{ij}^m \quad (1)$$
$$w_{ij} = z_i'\gamma_j + (\pi_j\sigma_i + \epsilon_{ij}^w) = z_i'\gamma_j + v_{ij}^w. \quad (2)$$

In equation (1), the dependent variable is the latent utility that individual i attaches to location j. Equation (1) is the utility if individual i chooses to stay in the home country. This utility depends on the wage at that location (w_{ij}), a vector of individual characteristics (x_i), and an individual and location-specific unobserved term ϵ^m. Notice that individuals differ in one unobservable characteristic (σ_i) that affects the utility of each alternative and that we term the propensity to migrate. The alternative-specific coefficient λ_j allows for differential effects on the propensity to migrate to each of the two destinations.[2]

Equation (2) specifies individual wages in each location as a function of observable (z_i) and unobservable characteristics (σ_i and ϵ_{ij}^w). Importantly, we allow for the propensity to migrate to affect also wages.[3] The following observation is both obvious and very important. Wages are only observed in one of the three locations for each individual. Thus, counterfactual wages need to be estimated. Regarding the stochastic specification, we assume that all random draws in

$$\{\epsilon_{i0}^m, \epsilon_{i1}^m, \epsilon_{i2}^m, \epsilon_{i0}^w, \epsilon_{i1}^w, \epsilon_{i2}^w, \sigma_i\} \quad (3)$$

are independent across individuals. Moreover, draws $\{\epsilon_{i0}^m, \epsilon_{i1}^m, \epsilon_{i2}^m\}$ are independently distributed across alternatives with c.d.f. F^m. Similarly, $\{\epsilon_{i0}^w, \epsilon_{i1}^w, \epsilon_{i2}^w\}$ are independently distributed across alternatives with c.d.f. F^w. The c.d.f. of propensity to migrate σ_i is F^σ. Naturally, we also assume that the covariate vectors (x_i and z_i) are uncorrelated with ϵ_{ij}^m and ϵ_{ij}^w.

Importantly, individual unobserved heterogeneity has two important implications. First, it causes the unobservable component in the migration equation (v_{ij}^m) to be correlated across

[2] We normalize λ_0. As a result, σ_i can be interpreted as the individual propensity to migrate. We note that if $\lambda_j > 0$ for both destinations ($j = 1, 2$), then there will be positive correlation between v_{i1}^m and v_{i2}^m.

[3] As with equation (1), we normalize π_0. Note also that if π_j and λ_j are positive for both destinations then high-σ_i individuals will be more likely to migrate and will also be more likely to obtain above-average earnings. However, this need not be the case for all groups of individuals, as defined by observable characteristics.

destinations, for a given individual. However, under our assumptions, there is no correlation between the error term in location 0 and in locations 1 or 2. Specifically,

$$E(v_{i1}^m v_{i2}^m) = \lambda_1 \lambda_2 E(\sigma_i^2) \tag{4}$$

$$E(v_{i0}^m v_{ij}^m) = 0 \qquad j = 1, 2 \tag{5}$$

Secondly, unobserved heterogeneity introduces a selection bias in the estimation of the wage equation. Namely, migrants are not a random sample of the total original population. Under some conditions, they will have above-average ability. One possible interpretation for σ_i in this case is that it is measure of each individual's degree of risk aversion. It is plausible that less risk averse individuals are more likely to migrate. At the same time, they are likely to self-select into more risky jobs. These jobs are likely to pay higher wages, compared to jobs requiring similar skills, to compensate for the higher risk. But it is also possible that migrants are positively selected (in unobservables) to one destination and negatively to another.

Let us briefly discuss the identification of this model. The wage regression is essentially a standard Mincer regression, where the coefficients are identified from individual variation in each location. Turning to the migration equation, the income coefficient (α) is identified from individual variation both within and across locations. In contrast, in studies using only macro data, the identification is purely based on the correlation between the proportion of migrants and average wages across destinations. Given that often wage data across countries is often not perfectly comparable, our identification of the income parameter is more appealing. Finally, it is well known that in random utility models (as ours), not all the coefficients on the individual-specific characteristics in the migration equation are identified. We follow the convention of normalizing $\beta_0 = 0$. The identification of the terms in the variance-covariance matrix of the migration equation depend on the stochastic specification of the model, which we discuss in the next section.

The coefficients of the migration equation have a direct interpretation in terms of utility. The coefficient on predicted income represents the marginal utility of income. The coefficients on the variables that are not location-specific are interpreted as relative with respect to the normalizing location (the origin country in this case). Thus, we interpret the variables of the migration equation other than income as reflecting the *net* gain from migration to a particular destination. Obviously, by changing the sign, these variables can be interpreted as the net cost from migrating to each destination.

3 Estimation

Typically, due to data constraints, the versions of the Roy model that are estimated in the literature on international migration flows are special cases of the model above. In particular, researchers are usually forced to assume that there is no relevant unobserved heterogeneity. Under that assumption, the model does not display correlation across alternatives among the error terms for a given individual. In that case, assuming F^m is the c.d.f of a type-one extreme value distribution, the model becomes the well-known conditional logit, which can be estimated easily by maximum likelihood.[4]

In contrast, the literature on internal migration has argued that unobserved heterogeneity cannot be ignored. The richer data available to study internal migration flows has led researchers to develop techniques to estimate versions of the Roy model where unobserved heterogeneity is taken into account (Dahl (2002), Kennan and Walker (2003) or Bayer, Khan, and Timmins (2008), among others). The key aspect of our paper is that we have individual level data for both migrants and non-migrants, covering all the relevant migration destinations for the episode at hand.

As noted in the previous section, unobserved heterogeneity introduces two challenges. First, note that the error terms in the migration equation are correlated across alternatives.[5] This is because, for example, high-σ_i individuals might be more likely to migrate. As a result, estimating a conditional logit model would not be appropriate, as it requires independent draws across alternatives. The particular structure of correlations in equations (4) and (5) nicely fits the structure of both the nested logit (MacFadden (1978) and the multinomial probit.

There is yet a second challenge in the estimation of the model: unobserved heterogeneity introduces a selection problem in the wage equation. As mentioned earlier, income is only observed in one of the three locations for each individual. Thus, their counterfactual income

[4]When only cross-country, macro data is available, researchers usually estimate the aggregate relationship that can be derived from the conditional logit setup(Grogger and Hanson (2008)).

[5]That is, unobserved heterogeneity implies that the property of independence of irrelevant alternatives (IIA) will not hold.

needs to be estimated in the other two locations. However, the observed sample of individuals in a given destination will, most likely, not be a random sample of the population of origin. Migrants will tend to have high migration propensities to that destination, which might be correlated with their earnings draws. As a result, the estimated coefficient for the marginal utility of income (α) is likely to be biased.

The estimation of our model can be divided into two stages. First, we estimate individual, location-specific wages correcting for self-selection using the method in Dahl (2002). Second, using those predictions, we estimate a nested logit and a multinomial probit model. These models take into account the discrete nature of the location choice and are able to accommodate the structure of correlations that unobserved heterogeneity introduces. We next provide some more details on our estimation method.

Let us start with the self-selection problem in the wage equation. We apply the correction proposed by Dahl (2002) in three steps. First, we divide the population into mutually exclusive cells defined by observable characteristics: age, education, gender, marital status, and household size. Second, for each cell, we compute the proportion of individuals that chose to stay in the home country ($j = 0$) and the proportions that chose to migrate to each destination ($j = 1, 2$). Third, we estimate the following selection-corrected earnings equation:

$$w_{ij} = z_i'\gamma_j + f_j(p_j) + \epsilon_{ij}^w \tag{6}$$

where the new term is a polynomial function of the proportion of individuals that choose location j. Intuitively, it corrects for the fact that migrants to a particular destination have a higher propensity to migrate to that destination, which is likely to influence their earnings as well. The assumption behind Dahl's method is that unobservable heterogeneity within cells is relatively small.[6]

Once we obtain the selection-corrected individual income in each location, we turn to the second stage, the discrete choice migration problem. As noted earlier, under appropriate

[6]Note that the correction term is indexed by j. Thus, here we are allowing for the degree of selection to vary by destination. The correction is more efficient when including a more general polynomial, containing also the proportions of individuals in the other locations. We follow this approach, as explained below. Bourguignon, Fournier, and Gurgand (2007) compare Dahl's estimation procedure with others previously developed and widely used by the literature: Lee (1983) and Dubin and McFadden (1984). They conclude that Dahl (2002) and their own variant of Dubin and McFadden (1984) are preferable to Lee's method.

distributional assumptions, our model becomes the nested logit or the trinomial probit. Both models capture the discrete nature of the migration choice (only three locations) and allow for correlated shocks across locations. The correlation pattern implied by these models is not fully unrestricted, but it matches well the correlation structure in our model. In the language of the nested logit model, the three locations can be partitioned in two nests: the trivial nest containing only the home country (0) and the nest of potential destinations (1 and 2). Both the nested logit and the trinomial probit allow for correlation within nests but require zero correlation across nests.[7] The main difference between the two empirical models is that the trinomial probit is more general, as it allows for heteroscedasticity and for negative correlation across alternatives. As a result, maximum likelihood estimation of the trinomial probit will be our preferred estimation.

4 The Ecuadorian crisis and migration

At the end of the 1990s, Ecuador was hit by a major economic and financial crisis, which can be traced back to an array of deeply-routed long-term economic and institutional factors (Beckerman (2002)), and which was triggered by a series of adverse shocks. Specifically, the price of oil, that constitutes the single largest revenue item in the Ecuadorian Balance of Payments and a crucial fiscal resource, reached a historical low in 1998, when it fell short of 8 dollars per barrel, contributing to a mounting fiscal and current account deficit. In the same year, the coastal provinces suffered from the floods induced by El Niño rains, which caused major infrastructure disruptions and severely hurt the agricultural sector, with a 2.6 billion dollars estimated total damage, representing 13 per cent of GDP (IMF (2000)).

These events compounded the macroeconomic instability of the country. They led to the collapse of the domestic currency, the Sucre, and to a large-scale banking crisis. The economic system was de facto dollarized at that time (Jácome (2004)), and domestic banks' balance sheets were seriously deteriorated by the depreciation of the sucre, which compromised the ability of domestic debtors to pay back the dollar denominated loans they had received. Their exposure to the export-oriented agricultural sector in the coastal provinces did not help either. Notwithstanding huge injections of liquidity on the side of the Central Bank and the introduction of a blanket public guarantee of all deposits, the fears of a widespread

[7] See Keane (1992) for more details.

banking crisis mounted, and the government froze all bank accounts in March 1999, in a desperate attempt to prevent a bank run. By the end of 1999, Ecuador was experiencing a 2-digit monthly rate of inflation, and its per capita real GDP had declined by 7.6 per cent over the year (World Bank (2008)). The government decided to adopt the dollar as a legal tender of exchange in January 2000, to avoid the incipient risk of hyperinflation, and to try to revive credit operations at a time when 16 out of 36 domestic banks had already been closed or had gone under public stewardship (Jácome (2004)).

Dollarization was implemented at a markedly undervalued conversion rate, as the decision to dollarize had not been agreed upon with any international financial institution, so that the Central Bank had to buy back the domestic monetary base with its limited holdings of foreign reserves. This entailed a dramatic reduction in real wages, and it prevented dollarization from immediately bringing the increase in prices to a halt: the consumer price index rose by 96 per cent in 2000, with a massive "once-and-for-all price-level adjustment" (Beckerman and Cortés-Douglas (2002)). As Figure 1 shows, Ecuador experienced a moderately positive rate of growth in per capita GDP in 2000, which strengthened since 2001, when the beneficial effects of dollarization on price stability began to appear.

(Figure 1)

Notwithstanding the economic recovery, which was mostly driven by the increase in the world price of oil and by the construction of a second Transandean pipeline increasing the country export capacity, the crisis had produced some long-lasting effects on Ecuadorian households. High inflation had substantially eroded the real value of their savings, as the unfreezing of banking deposits was completed only in March 2000 (Beckerman and Cortés-Douglas (2002)) and some payments to depositors in failed banks where still pending almost ten years after the crisis (Laeven and Valencia (2008)), although the government had extended a public blanket guarantee on all deposits in 1999.

The 1999 crisis triggered an unprecedented wave of international migration out of Ecuador, with approximately 600,000 individuals leaving over the 1999-2005 period (Ramírez Gallegos and Ramírez (2005)), from a country with a total population of 12 million, according to the 2001 national census. Flows were mostly directed to just two major destinations, namely the United States and Spain, which attracted between 80 to 90 per cent of the Ecuadorian migrants who left after the crisis[8].

[8]Estimates from the 2001 Ecuador Census (for the beginning of the crisis) and from the ENEMDU 2005

(Figure 2) (Figure 3)

Figures 2 and 3 show that for both US and Spain 1999 marked the beginning of this migration wave, while yearly flows had returned to almost their pre-crisis level for both countries by 2005. Although the timing of the migration episodes to both destinations is similar, the size of the flows is not: while the size of the Ecuadorian community in the US increased from 272,000 individuals before the crisis, according to the 2000 US Census, to 394,000 in 2005, according to the American Community Survey 2007, the Ecuadorian community grew from 76,000 individuals before the crisis, according to the 2001 Spanish Census, to 457,000 individuals in 2005, as reported by a Spanish administrative data source, the Registry or Padrón[9]. Thus, these figures suggest that the size of Ecuadorian migration to Spain was 3 to 4 times larger than migration to the United States between 1999 and 2005.

It is important to mention that two features of the recent wave of Ecuadorian migration which emerge from Figures 2 and 3, namely the sudden discontinuity induced by the crisis and the limited number of years over which flows reached unprecedented levels, have important implications with respect to the relationship between education and the prospect to migrate. The theoretical literature on migrants' self-selection regards educational decisions as unaffected by the prospect to migrate (e.g. Chiquiar and Hanson (2005); McKenzie and Rapoport (2009)), while the literature on the so-called brain gain maintains that individuals adjust their educational decisions in response to the possibility of migrating to a foreign high-wage country (e.g. Beine, Docquier, and Rapoport (2001). This argument may cast doubts on the assumption of exogeneity of the educational decisions that underlies the empirical tests of self-selection models, but these doubts can be safely dismissed as far as the recent Ecuadorian migration episode is concerned. The migration wave over the 1999-2005 period was triggered by a sudden economic crisis, so that Ecuadorian migrants had little or no time to adjust their educational decisions.

for the rest of the years.

[9]The Padrón would suggest that just 8,000 Ecuadorians resided in Spain in 1998, but this figure is unreliable as recent Ecuadorian migrants had little incentives to register; the situation changed in April 2000, when the government approved an amnesty which granted to all registered immigrants free access to health care and education, so that later figures have a reliability that prior ones do not have.

5 Data

5.1 Data sources

The estimation of our model requires to have individual-level data on the migrants which went to both the US and Spain over our reference period, and for the population of stayers in Ecuador. On the Spanish side, we rely on the Encuesta Nacional de Inmigrantes, ENI, which was performed in 2007 on a sample of 15,500 foreign born resident in Spain, and which represents the first nationally representative immigrant survey conducted in Spain. The timing of the Spanish survey fits well with the need to focus on the Ecuadorian migrants who left over the 1999-2005 period, as it is well known that immigrant surveys are unable to adequately enumerate recent migrants. On the US side, we resorted to the American Community Survey, ACS, conducted in 2007, whose sample covers approximately 2.5 per cent of the resident population, to extract data on recent Ecuadorian immigrants. As far as the Ecuadorian side is concerned, the choice fell on the Encuesta Nacional sobre el Empleo y Desempleo en el Area Urbana y Rural, ENEMDU, a nationally representative labor market survey that is conducted once a year. We chose the 2005 round of this survey, as the argument that surveys are able to adequately represent migrants only with a delay clearly does not apply to stayers. Furthermore, and most importantly, having a 7-year reference period entails that the characteristics of the population of stayers are not influenced by sizeable cohort-effects, so that the relevance of the choice of the specific round of the ENEMDU can be downplayed[10]

The three different datasets provide comparable information on a set of individual-level characteristics, like age, gender, marital status, education and years since migration, and information about the employment status, characteristics of the occupation, and the level of labor income. Our sample is obtained through a merge of the three datasets, limited to all the individuals who were born in Ecuador between 1949 and 1982, and (when they did) left Ecuador between 1999 and 2005, our reference period. These individuals were aged between 16 and 49 years old and living in Ecuador in 1998, at the onset of the recent migration episode, so that we focus on the age groups where migration is most likely to occur. We opted for restricting our sample to a specific set of years of birth than on to a set of ages at the time of the survey or at the time of migration, as the former choice ensures that we

[10]We actually performed several robustness checks on the 2001 ENEMDU, with very similar results.

compare migrants with stayers belonging to the same birth cohorts. Otherwise, we would have violated the well established fact that the probability to migrate varies substantially over the life cycle of an individual. Our final joint sample is represented by 30,190 stayers from the ENEMDU 2005, 509 migrants to the US from the ACS 2007, and 949 migrants to Spain from the ENI 2007[11]. As a robustness check, we also restrict our estimates to the sample of individuals aged between 25 and 49 in 1998, although such an alternative restriction on age halves our sample size, as it allows to downplay the possible concern that younger migrants completed their education at destination rather than at home.

5.2 Descriptive statistics

Table 1 presents some basic descriptive statistics on stayers and migrants to the two destinations for individuals aged 16 to 64 years old, that is, before restricting to our joint sample. Migration to the US over the 1999-2005 period was predominantly male, while the reverse occurred for Spain. The share of male college graduates to the US is higher than the corresponding share among stayers, but the observed difference lacks statistical significance, while the descriptive statistics point to a negative selection of male Ecuadorian migrants to Spain, where the share of college graduates is just 8 per cent[12]. A different picture emerges when looking at females, where migrants to both the US and Spain appear to be positively selected with respect to education, although the difference with the latter destination is not significant at conventional confidence levels.

(Table 1)

Table 1 also reports data on income, which is defined as gross labor earnings over the 12 months prior to the survey, and that it has been converted in 2005 US dollar terms using domestic consumer price indices for both Spain and the US[13]. The data evidence that the relative income gains from migration are larger for females than for males, as the gender wage gap, which can be observed in all the three countries, is substantially larger in Ecuador

[11]The figure for the US amounts to 70 per cent of all the Ecuadorian migrants sampled in the ACS 2007, while the corresponding figure is 77 per cent for the ENI 2007, in line with the markedly increased intensity of migration flows to both destinations over our reference period.

[12]Following a well-established convention, individuals are regarded as college graduates if they have at least 4 years of college.

[13]The ENEMDU 2005 has a monthly reference period for labor earnings, so that the data have been annualized to ensure comparability with the other two data sources.

than in either of the two destinations. Conversely, males enjoy on average a larger increase in income when moving to the US or to Spain than females do. Furthermore, Table 1 reveals that Ecuadorian migrants to the US earn roughly 60 per cent more than migrants to Spain, and this holds true for both genders. This huge income differential, which can be reduced but not offset by an adjustment for purchasing power parity, can be reconciled with the scale of Ecuadorian migration flows across the two foreign destinations only through a major role played by migration costs.

5.3 Returns to skill, selection and sorting across destinations

Table 2 concentrates on our sample of individuals living in Ecuador and aged 16 to 49 years old in 1998. The left panel of Table 2 compares the employment rates of Ecuadorians in the three countries, broken down by gender and education. Males experience similar nearly identical employment rates in the three countries, and no major differences by educational group stand out from the descriptive statistics. For females the picture is more divers. For college graduates, the employment rate in Spain and, to a lesser extent, the US stands above the one observed in Ecuador, while non-college graduates experience a similar employment rate in Ecuador and Spain, while this falls substantially in the US[14].

(Table 2)

The right panel of Table 2 provides a similar breakdown of median labor earnings by education and gender. The picture that emerges from the data with respect to Ecuador and the US is in line with Grogger and Hanson (2008), who argue that skill-related differences in wage levels are increasing with the income level of a country, while relative returns to skill decline with income. The ratio of the median wage of a college graduate to the wage of a non college graduate is close to 3 in Ecuador, while it is less than 1.5 in the US, while, for males, the difference in the level of wages are 24,492 and 19,160 dollars for college and non-college graduates respectively. The picture for Spain is different, as there is no skill premium for being a college graduate neither for males nor for females.

(Table 3)

[14]Table 2 also evidences that the employment rates of female Ecuadorian migrants to the US fall substantially short of the corresponding rate for Spain; this difference can be probably traced back to the larger share of tied movers among the women who moved to the US, given the longer history of male migration to the country (Jokisch and Pribilsky (2002)) which favored family reunification (Sánchez (2004)).

Table 3 shows that the observed pattern of skill premia survives once we control for some observables in a simple Mincerian regression, where we resort to an OLS estimation of log labor earnings which disregards the issues of selection into both employment and destination discussed in Section 2. For males, the skill premium in Ecuador stands at 82 per cent, 35 per cent in the US, and a statistically insignificant negative 6 per cent in Spain. The goodness of fit of the model, as far as Spain is concerned, is fairly poor, strengthening the observation from Table 3 that the distribution of labor earnings for Ecuadorian migrants to Spain is very compressed, not just across skill levels.

Table 4 provides some further descriptive evidence on the pattern of migrants' selection and sorting across destinations, as it reports the change in the probability of being a college graduate depending on the chosen country of residence from a simple probit model, where we control for a basic set of observables[15].

(Table 4)

As far as selection is concerned, the top panel of Table 4 shows that the positive selection of Ecuadorian male migrants to the US that emerged from Table 1 does not survive when controlling for some observables, while the reverse occurs for male migrants to Spain, as the reported negative change in probability is highly statistical significant. For female migrants, the descriptive evidence on selection provided in Table 4 is in line with what Table 1 evidenced. The bottom panel of Table 4 provides some additional evidence on migrants' sorting between the US and Spain, based on a similar probit regression that is run on the sample of Ecuadorian migrants alone: both males and females appear to be positively sorted towards the US, where the share of college graduates is significantly larger than the corresponding share in Spain even after controlling for differences in observables.

The preliminary evidence that we offer on the scale, selection and sorting of Ecuadorian migration can be compared with the predictions arising from the log and the linear utility models. In their simplest version, both fail as far as scale is concerned, as the observed income differentials between the US and Spain would have suggested that flows would have been predominantly directed towards the former destination, while the reverse occurred, as shown by Figures 2 and 3. Conversely, the evidence with respect to sorting is in line with the predictions of both specifications of an income-maximization model of migration

[15]This includes 5-year age groups, marital status - and the full set of their interactions, and the Ecuadorian province of origin.

decisions, which point to a positive sorting towards the US. The log and the linear utility model produce contrasting predictions as far as selection is concerned, as the log model would predict negative selection to both destinations based on the estimates of the skill premia from the set of Mincerian regressions. By the same token, the linear utility model would predict positive selection to the US, where the level difference in wages between college and non college graduates is larger than in Ecuador, and the opposite for Spain. The descriptive evidence provided in Tables 1 and 4 is different across genders, and depending on whether one controls or not for observables, so that the predictions from the two models can be neither rejected nor validated.

6 Implementation

The estimation of our model can be divided into two stages. The first requires producing individual-level estimates of income that correct for self-selection in migration. The second stage involves the estimation of a discrete-choice migration model that allows for correlation across alternatives. Let us now provide some detail on how we do this.

6.1 Selection-corrected individual earnings

Construction of cells We follow the methodology in Dahl (2002).[16] We build different cell structures for migrants and for stayers, to take into account their different sample sizes. For stayers, there are 48 different cells, defined by gender (male-female), education (college-non college), age (three age groups), marital status (married-non married) and household size (larger or smaller than two). The average cell size is 586 stayers with a maximum of 2,700 and a minimum of 9. For migrants, the limited sample size reduces the number of cells to 8, defined by gender, education and household size. The average cell size is 178 migrants with a minimum of 32 and a maximum of 386[17].

For each of these cells, the proportion of individuals who actually stay, migrate to the US or migrate to Spain will be used as the predicted probability that an individual belonging

[16] As a robustness check, we replicate our estimation procedure using other standard corrections for self-selection. The results do not change much. All these methods are available in Stata's package SELMLOG, developed by Bourguignon, Fournier, and Gurgand (2007).

[17] Using the same cell division for migrants and stayers (either the coarser with 8 cells or the finer with 48 cells) does not alter the results.

to the cell locates in each of the destinations in the following step.

Estimation earnings equation For the prediction of counterfactual incomes we slightly depart from Dahl's methodology in order to take into account another potential selection problem. Individuals first choose whether to migrate or not. Conditional on migration, they choose whether to work or not. We only observe earnings in a particular location for individuals who chose to migrate *and* to work. It is well known that immigrants tend to have very high employment rates, which is also the case in our particular episode. As a result, the bias due to selection into employment is likely to be small.[18] Nevertheless, we prefer to be coherent and deal explicitly with selection bias at all levels. Another advantage of our approach is that we treat our data more efficiently, by not having to discard non-working individuals in our estimation of the migration choice model, equation (1). Specifically, we replace the Mincer earnings regression in (2) by a Heckman (1979) system of equations (in employment and earnings). Naturally, to check robustness we also conduct the analysis using the more common and simpler methodology.

Specifically, for each individual i, compute the proportion of individuals that are like him in terms of observables (that is, that are in the same cell as she is) that chose location j. Denote this proportion by $\hat{p}_{ij}$. Next, we estimate jointly the system

$$\begin{aligned} \Pr\left(emp_{ij}=1\right) &= \Phi\left(z_i^{emp\prime}\gamma_j^{emp}+f_j\left(\hat{p}_{ij},\hat{p}_{ij'}\right)\right) && (7)\\ \ln w_{ij} &= z_i^{wage\prime}\gamma_j^{wage}+\rho_j\sigma_j\lambda\left(z_{ij}^{emp\prime}\gamma_j^{emp}+f_j\left(\hat{p}_{ij},\hat{p}_{ij'}\right)\right), && (8)\end{aligned}$$

where z_i^{emp} includes a constant, a college graduate dummy, a female dummy, age and its square, a marital status dummy and household size, which is the excluded variable from the earnings equation. Function $f_j\left(\hat{p}_{ij},\hat{p}_{ij'}\right)$ is Dahl's correction polynomial. The exact form is a second order polynomial in the retention probability for stayers and a second order polynomial in the retention and first-best probability for migrants plus an interaction term. Likewise, z_i^{wage} includes the same variables as z_i^{emp} except for household size. We also include occupational dummies (for ten occupational categories) in z_i^{wage}, which allows us to identify the marginal utility of income coefficient in the migration choice model without fully relying on non-linear functional forms.

[18]Dahl (2002) and, to our knowledge, all other authors ignore this issue.

Prediction of individual income in all locations Next, we predict income in all three destinations for all individuals in the sample, using the estimated equation 8.[19]

There are small but significant differences when predicting incomes with or without controlling for selection. Dahl's selection polynomial is significant in equation 7 for the US and for Ecuador but not for Spain. As for Heckman's ρ_j, it is negatively significant for both the US (-0.80) and Ecuador (-0.82) and positive but statistically insignificant for Spain (0.04). Comparing average predicted incomes with and without controlling for selection, we find that the correction makes US predicted income fall by 0.87 per cent, which points into the direction of positive selection to the US in terms of unobservables. Spain's predicted income increases on average by 1.33 per cent, which would mean negative selection on unobservables. Finally, Ecuador's predicted income falls by 0.88 per cent, a sign that migrants are on average negatively selected on unobservables, which is consistent with the US and Spain evidence since most Ecuadorians (approximately three quarters) in our sample migrated to Spain.

6.2 The migration choice

Using the predicted incomes from 8, we estimate a discrete choice migration model that allows for correlation across alternatives. We control for all the variables that were used in the previous steps: gender, education, age, marital status and household size. As noted earlier, we experiment with two stochastic specifications: the nested logit and the more general multinomial (trinomial) probit. Following up on the discussion on Grogger and Hanson (2008) on whether income or the log of income is a more appropriate specification, we estimate both and compare the results.

[19]We also experiment with an alternative strategy to predict individual income. Namely, one can use actual income for the location that has been observed for each individual and then use predictions for the two counterfactual locations only. This does not play a major role in our main estimates. The results from all of these auxiliary regressions are available from the authors upon request.

7 Estimation results[20]

This section presents the main results of the paper. Using individual-level data from three different countries, we estimate a version of the Roy model that allows for observed and unobserved heterogeneity (both in migration propensities and in earnings). Unobserved heterogeneity introduces correlation across locations for a given individual (violating the IIA assumption) and a selection bias in the earnings equation.

First, we test the Independence of Irrelevant Alternatives (IIA) hypothesis. After that, we present our estimates of the model that allows for unobserved heterogeneity.

7.1 Testing the IIA hypothesis

It is common in the empirical literature that estimates versions of the Roy model based on aggregate flows data to employ empirical models that assume the Independence of Irrelevant Alternatives (IIA) property. That is, the odds of going to one location with respect to another are not affected by any third alternative. It is very simple to check whether the IIA assumption holds in our case. We can run the conditional logit (which assumes IIA), on the whole model and then run restricted models dropping one destination at a time. The odds of migrating to Spain with respect to staying in Ecuador should not be affected by the inclusion or exclusion of the US in the specification.

This exercise is reported in Table 5. For the model where income enters linearly, we see that the estimated coefficient of income varies widely across models, ranging between 0.064 and 0.560. For the model with log income, this coefficient varies much less across models, ranging between 0.791 and 0.916. However, our estimates are precise enough that we can reject the hypothesis that they are equal, both in the linear and the log case, using the Hausman test.[21]

[20]The standard errors of the estimations of the remaining sections have yet to be corrected for the sampling variability introduced by the selection step. We are confident that our results will continue to be statistically significant after this correction is made, assuming the standard errors increase twice as much as they increased for Dahl (2002) or even by 200 per cent since we have a double selection mechanism.

[21]Moreover, the estimates of the other parameters do vary substantially across models. The statistics at the bottom of the table show that we clearly reject the IIA hypothesis. We also reject the IIA hypothesis comparing the trinomial probit estimates to the appropriate binomial probits. Results available from the authors upon request.

In conclusion, it seems crucial to allow for correlated shocks across alternatives when estimating the migration choice equation. In the context of our model, this suggests that individual unobserved heterogeneity is also relevant when analyzing international migration flows.

7.2 Estimates

As noted earlier, our more flexible model is the trinomial probit, although we also estimate a nested logit for comparison. Table 6 presents the results, which also include the conditional logit for the sake of comparison.

Let us focus on the estimate of the income coefficient in the linear income model. Ignoring correlation across alternatives (the conditional logit), the income coefficient we obtain is 0.134. Allowing for a rich pattern of correlations (the trinomial probit) delivers a substantially lower coefficient equal to 0.038. Similarly, when we compare the estimates on income for the log income model, the trinomial probit delivers a lower estimate than the conditional logit. In both cases, accounting for correlated shocks reduces substantially the estimated effect of income.

It is also interesting to compare the estimates of the trinomial probit to those of the more restrictive nested logit. Both for the log and linear income models, the nested logit delivers much higher coefficients. We also note that the estimates of the nested logit do not appear robust to changes in the specification. While the log income model delivers a precisely estimated dissimilarity coefficient (equal to 0.492), the linear model produces a very high and very imprecise estimate for this coefficient (64.36), which is incompatible with the random utility interpretation of the nested logit.[22] For this reason, and its relatively restrictive assumptions, we strongly prefer the estimates produced by the trinomial probit.

It is interesting to compare our estimates of the income coefficient to those obtained using only aggregate data and, thus, under the assumption of no correlation on unobserved heterogeneity at the individual level. Grogger and Hanson (2008) estimate a cross-country regression that is consistent with the conditional logit model. Ortega and Peri (2009) perform a similar exercise but their specification is consistent both with the conditional logit and with

[22] Daly and Zachary (1979) and MacFadden (1978). Börsch-Supan (1990), Herriges and Kling (1996), Koning and Ridder (2003) and Ibáñez (2006), among others, establish different sets of more or less stringent consistency conditions. The [0,1) interval remains a relevant reference in most of them.

the nested logit, that is, it is correct even if there are correlated shocks across alternatives. Their specification still does not control for the selection bias in the earnings equation.

Our estimates for the effect of income in determining migration choices are fairly similar to the ones found in these two studies. In the specification with linear income, Grogger and Hanson (2008) obtain estimates that range between 0.018 and 0.103 and Ortega and Peri's estimate is around 0.060. In comparison, our preferred trinomial probit with linear income delivers 0.038. Turning to the log income model, Ortega and Peri (2009) find an estimate of 0.290, slightly lower than our 0.416.[23]

To sum up, we find that international migration flows are sensitive to earnings differences between the origin and destination country. Our estimates are remarkably similar in size to those found in some previous studies, even though we use earnings data and those studies use income per capita. Our estimates are robust to specification changes, such as log versus linear income, and are substantially lower when we allow for correlated shocks than when we do not.

8 The role of migration costs

As emphasized in Hanson (2008), the largest gap in the literature on international migration flows is a better understanding of the contribution of the different costs associated to migration in order to explain the pattern of migration flows in the data.[24]

The goal of this section is twofold. First, we use our estimates to compare the overall migration costs that Ecuadorian migrants faced when choosing whether to migrate to the US or to Spain. In doing so, we compare average migration costs across groups defined by gender and education. Secondly, we speculate which dimension of migration costs may have played a larger role in the migration episode that we have studied. We emphasize that our interpretation will only be based on suggestive evidence. Only because of the lack of studies on the nature and effects of migration costs and the difficulty in quantifying these effects do we dare to speculate on the basis of our limited evidence.

The vector of control variables in equation 1, along with the corresponding estimated coefficients, can be used to provide a measure of the net attractiveness of one destination

[23]In a similar specification to ours, Grogger and Hanson's estimate is large and negative.

[24]Ortega and Peri (2009) take on step in this direction by providing an estimate of the effects of a tightening of immigration policy in a country on the size of its inflows.

relative to the other. Obviously, this is also informative as to the size of the overall net migration costs associated to migration to one destination relative to the other.

Using the estimated coefficient (from our trinomial probit) on the marginal utility of income, these migration costs measured in "utils" can be transformed into 2005 US Dollars. We find it useful to report average migration costs, by destination, for different groups of individuals, indexed by g. These groups are defined in terms of gender and education (college versus non-college graduates). We report on the migration costs associated to the models where income enters linearly and in logs. That is,

$$\text{Linear utility model} \quad : \quad migcost_{gj}^{lin} = -\frac{\overline{x_g^{m\prime}}\hat{\beta}_j^m}{\hat{\alpha}} \tag{9}$$

$$\text{Log utility model} \quad : \quad migcost_{gj}^{\log} = -\frac{\overline{x_i^{m\prime}}\hat{\beta}_j^m}{\hat{\alpha}}\hat{w}_{gj} \tag{10}$$

Table 4 reports on the results of this exercise.[25] The top panel presents the cost implied by our estimates in the model with linear income and the bottom for the model with log income. Several features stand out. First, overall migration costs are much larger to the US than to Spain. Respectively, around $100,000 and below $20,000. This was to be expected given that average earnings were higher for all gender-education groups in the US than in Spain while the size of the flows from Ecuador to Spain was substantially larger than to the US. Secondly, for each given destination and controlling for education, migration costs appear to be slightly to moderately larger for males than for females. Third, migration costs seem to be slightly higher for non-college migrants, conditional on destination and gender. Finally, we note that the estimates arising from the log model are larger, in dollar terms, than those obtained from the linear model. In addition, the log model delivers more noticeable differences across the four groups.[26]

Our migration cost estimates are much larger than the available estimates of direct monetary costs that Ecuadorians incurred over this period in migrating to the US and to Spain. As suggested by our Figures 2 and 3, a sizeable fraction of Ecuadorian migration to the US and to Spain in this period is likely to have been undocumented. Informal estimates of the cost of illegally entering the US from Ecuador range between 7,000 and 9,000 dollars, which

[25]Keep in mind that we have normalized the migration costs of staying in Ecuador to zero.

[26]The log model predicts migration costs for males to the US increasing in the education level, in contrast to other gender-destination pairs and in contrast to the classical finding in the internal migration literature.

is at least ten times smaller than our estimates. In the case of Spain, we have direct evidence from the ENI. Surveyed Ecuadorians answered that the average cost of their trip to Spain was between 2,500 and 3,000 dollars, about one fifth of our estimate. In our view, the most plausible interpretation for the nature of the difference in estimated migration costs to the US and to Spain is the risk of apprehension. Over the period 1999-2005, the US had already stepped up enforcement. In contrast, migration to Spain (from Ecuador) at the beginning of the period was very easy and inexpensive. Moreover, the risk of apprehension was relatively low. As a result, the overall cost from migrating (illegally) to Spain is likely to have been very low, compared to the US.

Of course, our overall measures of migration costs also include other variables that enter into the migrants' calculations. Namely, the existence of ethnic networks, cultural and language distance, and access to public services. One can argue that the presence of Ecuadorians in the US prior to the 1999 crisis was a cost advantage relative to Spain, where Ecuadorian presence prior to the crisis was lower. On the other hand, cultural and linguistic distance and availability of public services to Ecuadorian migrants certainly contributed to make Spain a more attractive destination. We note that, by and large, the roles played by these three potentially important components of migration costs were constant over the period 1999-2005, which makes identification difficult.

There have been, however, some important immigration policy decisions in Spain during this period that can be considered as changes in the gap between the US and Spain's policy regarding the associated cost to migrating illegally. As noted earlier, around 1999 immigration policy toward illegal immigration was much tougher in the US than in Spain.[27] Moreover,

[27] Here is a daring attempt to quantify the larger probability of apprehension in the US, relative to Spain, during the period we are concerned about. According to Spain's Ministry of Work and Immigration, 15,000 Ecuadorian immigrants were returned to their country during the 1999-2003 period, the only one for which these data were available. Among these, 9 per cent were expulsions (immigrants already in Spain), 3 per cent devolutions and 88 per cent were returns (people rejected at official borders). The source is http://extranjeros.mtas.es. In comparison, the US Coast Guard intercepted 7,000 Ecuadorians in the Eastern Pacific during the 1999-2005 period (http://www.uscg.mil/hq/cg5/cg531/amio.asp#Statistics). In addition, the Mexican border patrol rejected almost 8,000 Ecuadorians between 2001 and 2005 who could arguably be in route to the United States (Instituto Nacional de Migración: http://www.inm.gob.mx/). If we extrapolate these numbers for the whole 1999-2005 period, it can be estimated that 21,000 illegal attempts of entry at Spain failed whereas 18,000 attempts of entry to the US failed. Taking into consideration that the flow of migrants to Spain was about three times larger than the flow to the US, these numbers point in

between 1999 and 2005 Spain conducted three regularization programs ("amnesties"), in 2000, 2001 and 2005. Figure 5 plots the inflows of Ecuadorians into Spain by month of entry.[28]

(Figure 5)

The most striking piece of evidence of the effects of changes in Spanish immigration policy on the size of the inflows from Ecuador is the following. Starting in August 2003, Ecuadorians were required to obtain a visa to enter Spain. Before that date, there was no visa requirement. As can be seen in Figure 5, monthly inflows dropped sharply between August and September 2003, remaining quite low after that. Economic conditions in Ecuador around this time were already improving but there was no sharp economic recovery that took place in the course of August 2003. In our view, the introduction of the visa requirement led to a sharp increase in the overall costs for Ecuadorians associated to migrating to Spain.[29]

Grogger and Hanson (2008) also compute relative migration costs for a group of countries according to the linear utility model. Their table 9 does not show Ecuador but it establishes that migration costs from Colombia to the US are 35,000 dollars larger than migration costs from Mexico to the United States. Their results could thus in principle be comparable to ours. However, when we replicate Grogger and Hanson's methodology in our data, by calculating their size, selection and sorting equations on our sample data and saturating the models with education and gender dummies to exactly identify the income coefficients[30], we

the direction of a more restrictive US immigration policy.

[28]The 2000 regularization included a provision that guaranteed access to the Spanish Health System (free health) and to the Spanish education system (free education for children) even for undocumented migrants as long as they register in the Padrón Municipal (Local Population Registry). Thus, since 2000, there is a very large incentive to register, making these data a very reliable source to count both documented and undocumented immigration.

[29]Apparently, there was a large excess demand for visas, which left many Ecuadorians unable to migrate to Spain.

[30]We take as predicted income the employment probability times the median income reported in table 2 for gender-education groups. We then run the following equation:

$\ln \frac{n_j^g}{n_{ec}^g} = \alpha \left(f\left(\hat{w}_j^g\right) - f\left(\hat{w}_{ec}^g\right)\right) + D_j^f + D_j^e + D^{fe}$

where n_j^g is the number of migrants in education-group g that went to country j; n_{ec}^g is the number of stayers in the same group; $f\left(\hat{w}_j^g\right) - f\left(\hat{w}_{ec}^g\right)$ is the wage differential between origin and destination for each group; D_j^f are destination-gender dummies; D_j^e are destination-education dummies; and D^{fe} is a gender-education dummy. There are eight observations (two destinations times four gender-education groups) and eight parameters to that allow an exact identification of α with a very flexible structure of migration costs.

find very different results in terms of migration costs. Concentrating on the linear utility model[31], the comparison can be seen in figure 6.

(Figure 6)

The macro estimates are especially inflated for the case of Spain even though the estimated income coefficients are actually quite similar (0.044 in the macro model versus 0.038 in the micro model). In conclusion, it seems that both controlling for selection and correcting the violation of the IIA assumption has a dramatic impact in the estimated migration costs from an utility maximization model.

9 Conclusions

This paper has presented a logically consistent approach to the estimation of an international migration model which controls for selection into migration due to unobserved individual-specific factors.

The main methodological contribution of this paper is to apply to the analysis of a recent large international migration episode, namely the Ecuadorian migration which followed the 1999 economic crisis, the empirical framework that is the workhorse of the empirical papers on the determinants of internal migration. The notably demanding data requirements that such an approach entails have been met merging individual-level datasets collected in Ecuador and in the two destination countries: Spain and the US, which attracted more than 80 per cent of the recent Ecuadorian migrants.

The estimation extends to the realm of international migration studies the well-established finding in the internal migration literature that failing to correct for the non-random selection of migrants would represent a serious source of bias.

We find that international migration decisions do respond to income differentials across possible destinations, even in a context where destination countries adopt markedly different immigration policies, and migration decisions were severely credit-constrained because of the crisis. The income-responsiveness of the decision to migrate is found independently on whether income is assumed to enter the utility function in a linear or logarithmic form. This suggests that an individual-level approach may end up validating both competing specifications, provided that unobserved ability and a flexible structure of migration costs are

[31] The results for the log utility model are very similar and can be requested from the authors.

consistently included in the estimation. Future tests of alternative migration-decisions models would have a greater cutting power if they were based on the analysis of the determinants of moving costs, something that, with some notable exceptions, has been so far just poorly explained.

References

Bayer, P., N. Keohane, and C. Timmins (2008): "Migration and Hedonic Valuation: The Case of Air Quality," *Journal of Environmental Economics and Management*, forthcoming.

Bayer, P., S. Khan, and C. Timmins (2008): "Nonparametric Identification and Estimation in a Generalized Roy Model," *NBER Working Paper Series*, 13949.

Beckerman, P. (2002): "Longer-term Origins of Ecuador's "Predollarization" crisis," *In Crisis and Dollarization in Ecuador, edited by Beckerman, P. and Solimano, A.*

Beckerman, P., and H. Cortés-Douglas (2002): "Ecuador under Dollarization: Opportunities and Risks," *In Crisis and Dollarization in Ecuador, edited by Beckerman, P. and Solimano, A.*

Beine, M., F. Docquier, and H. Rapoport (2001): "Brain drain and economic growth: theory and evidence," *Journal of Development Economics*, 64, 275–289.

Belot, M. V., and T. J. Hatton (2008): "Immigrant Selection in the OECD," *The Australian National University Centre for Economic Policy Research Discussion Paper Series*, 571.

Bertoli, S. (2009): "Networks, sorting and self-selection of Ecuadorian migrants," *Mimeo.*

Börsch-Supan, A. (1990): "On the compatibility of nested logit models with utility maximization," *Journal of Econometrics*, 43, 373–388.

Bourguignon, F., M. Fournier, and M. Gurgand (2007): "Selection Bias Correction Based on the Multinomial Logit Model: Monte Carlo Comparions," *Journal of Economic Surveys*, 21(1), 174–205.

Chiquiar, D., and G. H. Hanson (2005): "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2), 239–281.

Clark, X., T. J. Hatton, and J. G. Williamson (2007): "Explaining U.S. Immigration, 1971-1998," *The Review of Economics and Statistics*, 89(2), 359–373.

Clemens, M., C. E. Montenegro, and L. Pritchett (2008): "The Place Premium: Wage Differences for Identical Workers across the US Border," *Center for Global Development Working Paper*, 148.

Dahl, G. B. (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367–2420.

Daly, A. J., and S. Zachary (1979): "Improved multiple choice models," *In Identifying and measuring the determinants of mode choice, edited by D. Hensher and Q. Dalvi*, Teakfield, London, 335–357.

Dubin, J. A., and D. L. McFadden (1984): "An econometric analysis of residential electric appliance holdings and consumption," *Econometrica*, 52, 345–362.

Falaris, E. F. (1987): "A Nested Logit Migration Model with Selectivity," *International Economic Review*, 28(2), 429–443.

——— (1988): "Migration and Wages of Young Men," *The Journal of Human Resources*, 23(4), 514–534.

Gratton, B. (2007): "Ecuadorians in the United States and Spain: History, Gender and Niche Formation," *Journal of Ethnic and Migration Studies*, 33(4), 581–599.

Grogger, J., and G. H. Hanson (2008): "Income Maximization and the Selection and Sorting of International Migrants," *NBER Working Paper Series*, 13821.

Hanson, G. H. (2008): "International Migration and Development," *Commission on Growth and Development Working Paper*, 42.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Herriges, J. A., and C. L. Kling (1996): “Testing the Consistency of Nested Logit Models with Utility Maximization,” *Economics Letters*, 50, 33–39.

Ibáñez, J. N. (2006): “Consistency of Nested Logit Models with Utility Maximization,” *European Transport Conference.*

IMF (2000): “Ecuador: Selected Issues and Statistical Annex,” *IMF Staff Country Report*, 00/125.

Jácome, L. I. (2004): “The Late 1990s Financial Crisis in Ecuador: Institutional Weaknesses, Fiscal Rigidities, and Financial Dollarization at Work,” *IMF Working Paper*, 04/12.

Jokisch, B., and J. Pribilsky (2002): “he Panic to Leave: Economic Crisis and the ”New Emigration” from Ecuador,” *International Migration*, 40(4), 76–101.

Keane, M. P. (1992): “A Note on Identification in the Multinomial Probit Model,” *Journal of Business and Economic Statistics*, 10(2), 193–200.

Kennan, J., and J. R. Walker (2003): “The Effect of Expected Income on Individual Migration Decisions,” *NBER Working Paper Series*, 9585.

Koning, R. H., and G. Ridder (2003): “Discrete choice and stochastic utility maximization,” *Econometrics Journal*, 6, 1–27.

Laeven, L., and F. Valencia (2008): “Systemic Banking Crises: A New Database,” *IMF Working Paper*, 08/224.

Lee, L.-F. (1983): “Generalized Econometric Models with Selectivity,” *Econometrica*, 51, 507–512.

MacFadden, D. (1978): “Modelling The Choice of Residential Location,” *In Spatial Interaction Theories and Models, edited by Karlgvist, A. et al.*

Mayda, A. M. (2008): “International migration: A panel data analysis of the determinants of bilateral flows,” *Journal of Population Economics*, forthcoming.

McKenzie, D. J., and H. Rapoport (2009): "Self-selection patterns in Mexico-U.S. migration: The role of migration networks," *The Review of Economics and Statistics*, forthcoming.

Nakosteen, R. A., and M. A. Zimmer (1980): "Migration and Income: The Question of Self-Selection," *Southern Economic Journal*, 46(3), 840–851.

Ortega, F., and G. Peri (2009): "The Causes and Effects of International Labor Mobility: Evidence from OECD Countries 1980-2005," *NBER Working Paper Series*, 14833.

Ramírez Gallegos, F., and J. Ramírez (2005): "La Estampida Migratoria Ecuatoriana: Crisis, redes transnacionales y repertorios de acción migratoria," *Abya-Yala, Quito.*

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

Ruggles, S., M. Sobek, T. Alexander, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander (2008): *Integrated Public Use Microdata Series: Version 4.0 [Machine-readable database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor]. http://usa.ipums.org/usa/.

Sánchez, J. (2004): "Ensayo sobre la economia de la emigración en Ecuador," *Ecuador Debate*, 63, 47–62.

World Bank (2008): "World Development Indicators," Washington.

TABLES

Table 1: Descriptive Statistics

All individuals 16-64	Stayers			United States			Spain		
	mean	std.dev.	s.e	mean	std.dev.	s.e	mean	std.dev.	s.e
female, share	0.51	0.50	0.00	0.45	0.50	0.02	0.54	0.50	0.02
Males									
Age at migration	35.02	13.61	0.12	26.92	9.51	0.56	27.70	8.95	0.54
years since migration				5.30	2.08	0.13	6.12	1.42	0.08
college grad., share	0.11	0.31	0.00	0.12	0.32	0.02	0.08	0.27	0.01
Labor income, 2005 USD	2,873	4,926	52	25,056	18,774	1,312	15,728	4,454	272
Females									
Age at migration	35.39	13.34	0.12	28.92	10.31	0.62	27.52	9.02	0.58
years since migration	0.00	0.00	0.00	5.34	1.96	0.13	5.82	1.44	0.10
college grad., share	0.10	0.30	0.00	0.19	0.39	0.03	0.13	0.34	0.02
Labor income, 2005 USD	1,213	2,815	30	17,071	12,074	1,206	10,567	3,630	222
obs		41355			662			1124	
Source		ENEMDU			ACS			ENI	
Year		2005			2007			2007	

Table 2: Employment rates and median earnings in 2005 US Dollars for individuals aged 16 to 49 in 1998

Ecuadorians aged 16 to 49 in 1998

	EMPLOYMENT RATES			MEDIAN EARNINGS IN 2005 DOLLARS		
	Ecuador	US	Spain	Ecuador	US	Spain
College Graduates						
men	0.92	0.92	0.92	6,000	30,492	15,431
women	0.80	0.63	0.84	4,344	20,582	10,942
Non College Graduates						
men	0.94	0.90	0.90	2,280	21,440	15,431
women	0.54	0.63	0.79	1,560	14,865	10,521

Table 3: Skill Premia

Mincer Regressions for individuals aged 16 to 49 in 1998

Dependent variable:	Ecuador (ENEMDU 2005)		US (ACS 2007)		Spain (ENI 2007)	
log income in 2005	men	women	men	women	men	women
College Graduate	0.824	0.959	0.346	0.332	-0.062	-0.023
	[0.027]***	[0.029]***	[0.109]***	[0.122]***	[0.049]	[0.049]
Age	0.042	0.002	-0.039	0.007	-0.008	0.003
	[0.009]***	[0.014]	[0.049]	[0.049]	[0.017]	[0.028]
Age sq.	0.000	0.000	0.000	0.000	0.000	0.000
	[0.000]***	[0.000]	[0.001]	[0.001]	[0.000]	[0.000]
			0.053	0.025	0.018	0.003
Years since migration			[0.020]***	[0.029]	[0.010]*	[0.015]
Constant	6.888	7.260	10.458	9.347	9.678	9.069
	[0.174]***	[0.266]***	[0.949]***	[0.886]***	[0.332]***	[0.500]***
Observations	10,679	6,353	188	136	378	373
R-squared	0.15	0.18	0.08	0.05	0.02	0.02

legend: * p<.1; ** p<.05; *** p<.01

Table 4: Selection and sorting (probits on being a collage graduate)

Probits on the share of college graduates

Effects of discrete changes reported	Males aged 16 to 49 in 1998				
Selection to the United States	-0.002	0.011	0.01	0.01	
	[0.022]	[0.024]	[0.024]	[0.024]	
Selection to Spain	-0.066	-0.057	-0.059	-0.054	-0.052
	[0.015]***	[0.016]***	[0.015]***	[0.016]***	[0.016]***
Age groups	No	Yes	No	Yes	Yes
Year of birth dummy	No	No	Yes	No	No
Marital status	No	No	No	Yes	Yes
Province of origin	No	No	No	No	Yes
Observations	15,294	15,294	15,294	15,294	15,035
	Females aged 16 to 49 in 1998				
Selection to the United States	0.083	0.081	0.084	0.093	
	[0.030]***	[0.030]***	[0.030]***	[0.031]***	
Selection to Spain	0.019	0.012	0.012	0.02	0.018
	[0.021]	[0.021]	[0.021]	[0.021]	[0.021]
Age groups	No	Yes	No	Yes	Yes
Year of birth dummy	No	No	Yes	No	No
Marital status	No	No	No	Yes	Yes
Province of origin	No	No	No	No	Yes
Observations	16,320	16,320	16,320	16,320	16,047

	Males aged 16 to 49 in 1998			
Sorting to Spain	-0.064	-0.063	-0.072	-0.065
	[0.027]**	[0.026]**	[0.025]***	[0.026]***
Age groups	No	Yes	No	Yes
Year of birth dummy	No	No	Yes	No
Marital status	No	No	No	Yes
Province of origin	No	No	No	No
Observations	688	688	688	688
	Females aged 16 to 49 in 1998			
Sorting to Spain	-0.065	-0.064	-0.068	-0.06
	[0.036]*	[0.035]*	[0.035]**	[0.034]*
Age groups	No	Yes	No	Yes
Year of birth dummy	No	No	Yes	No
Marital status	No	No	No	Yes
Province of origin	No	No	No	No
Observations	736	736	736	736

Table 5:

Probability of choosing a location (conditional logit model)

	All countries		Dropping the US		Dropping Spain	
Linear income	0.134***		0.560***		0.064***	
Log income		0.885***		0.916***		0.791***
United States						
College	-0.975***	0.152			-0.494***	0.108
Female	0.634***	0.194*			0.227*	0.181
Age	0.060	0.072			0.085	0.085
Age sq.	-0.001	-0.002**			-0.002**	-0.002**
Married	0.838***	1.069***			0.917***	1.069***
Household size	-0.564***	-0.582***			-0.586***	-0.596***
Constant	-5.323***	-4.664***			-4.166***	-4.629***
Spain						
College	-0.225	-0.015	0.523***	-0.003		
Female	0.569***	0.395***	2.548***	0.404***		
Age	0.305***	0.299***	0.277***	0.305***		
Age sq.	-0.005***	-0.005***	-0.005***	-0.005***		
Married	0.459***	0.521***	0.838***	0.524***		
Household size	-0.516***	-0.522***	-0.528***	-0.532***		
Constant	-7.136***	-7.059***	-11.745***	-7.174***		
Log-likelihood	-1256681.8	-1258053.2	-852298.4	-874359.0	-373071.2	-373328.48
Observations	29546		29037		28631	
Hausman Test						
Chi-2			267.46	39.88	100.22	25.86
p-value			0.0000	0.0000	0.0000	0.0011

legend: * p<.1; ** p<.05; *** p<.01

Table 6:

Probability of choosing a location

	Conditional Logit		Nested Logit		Trinomial Probit	
Linear income	0.134***		0.308***		0.038***	
Log income		0.885***		0.821***		0.416***
United States						
College	-0.975***	0.152	-1.093	0.015	-0.219**	0.101
Female	0.634***	0.194*	-3.424	0.216**	0.130*	0.09
Age	0.060	0.072	-3.581	0.155***	0.02	0.019
Age sq.	-0.001	-0.002**	0.049	-0.003***	-0.001	-0.001
Married	0.838***	1.069***	7.658	0.826***	0.517***	0.596***
Household size	-0.564***	-0.582***	2.778	-0.562***	-0.270***	-0.275***
Constant	-5.323***	-4.664***	-54.422	-5.230***	-2.863***	-2.997***
Spain						
College	-0.225	-0.015	-0.081	-0.011	0.035***	0.003
Female	0.569***	0.395***	2.844	0.345***	0.171***	0.160***
Age	0.305***	0.299***	1.371	0.269***	0.018***	0.118***
Age sq.	-0.005***	-0.005***	-0.02	-0.004***	-0.000***	-0.002***
Married	0.459***	0.521***	-1.252	0.583***	0.057***	0.221***
Household size	-0.516***	-0.522***	-1.523	-0.527***	-0.035***	-0.217***
Constant	-7.136***	-7.059***	-39.735	-6.282***	-0.827***	-3.213***
Log-likelihood	-1256681.8	-1258053.2	-1238554.4	-1257406.6	-1239093.0	-1262545.9
Observations			29546			
Dissimilarity coefficient			64.363	0.492***		
US-Spain correlation	0.00	0.00	-4141.60	0.76	-0.80	-0.76

legend: * p<.1; ** p<.05; *** p<.01

FIGURES

Figure 1: Macroeconomic Conditions in Ecuador (1995-2005)

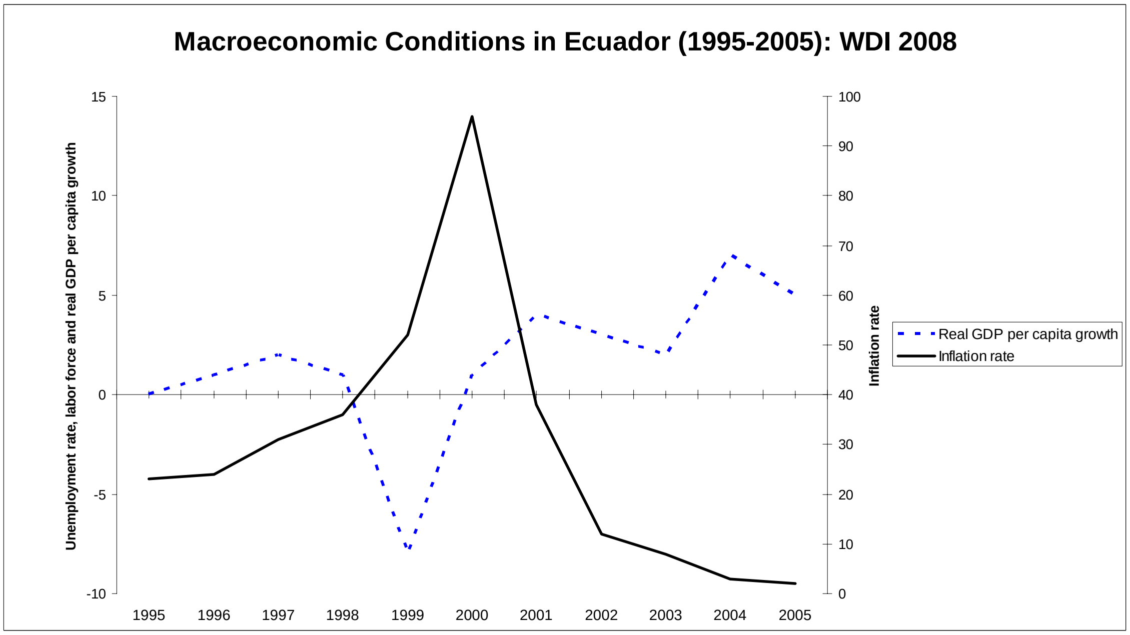


Figure 2: Arrivals of Ecuadorians in the US according to the ACS 2007

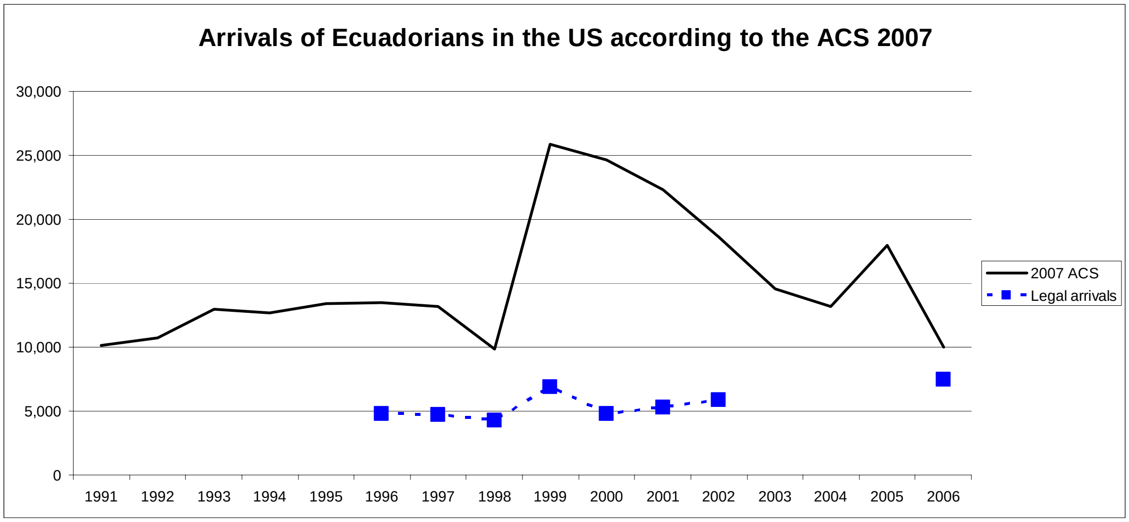


Figure 3: Arrivals of Ecuadorians in Spain according to the ENI 2007

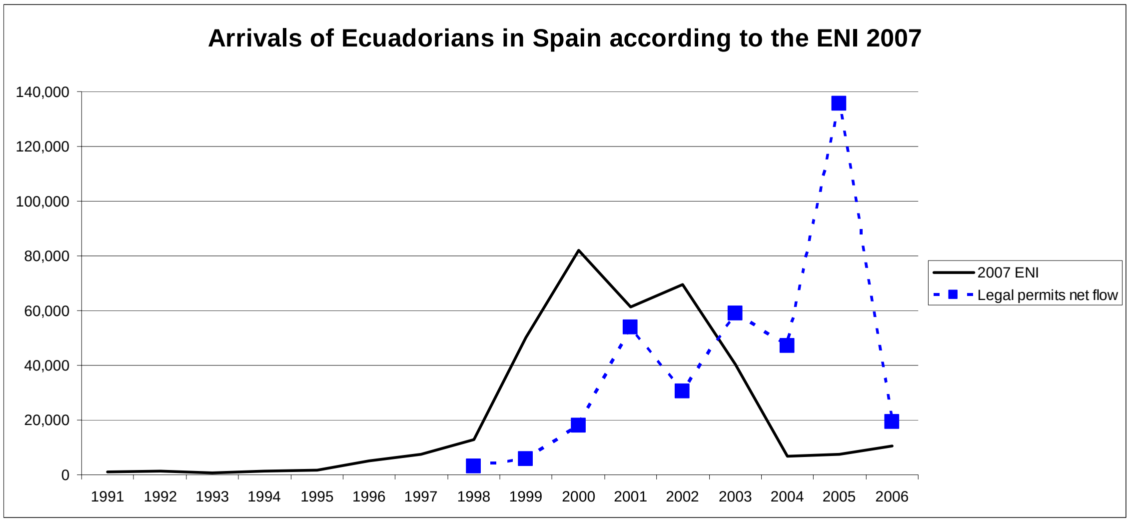
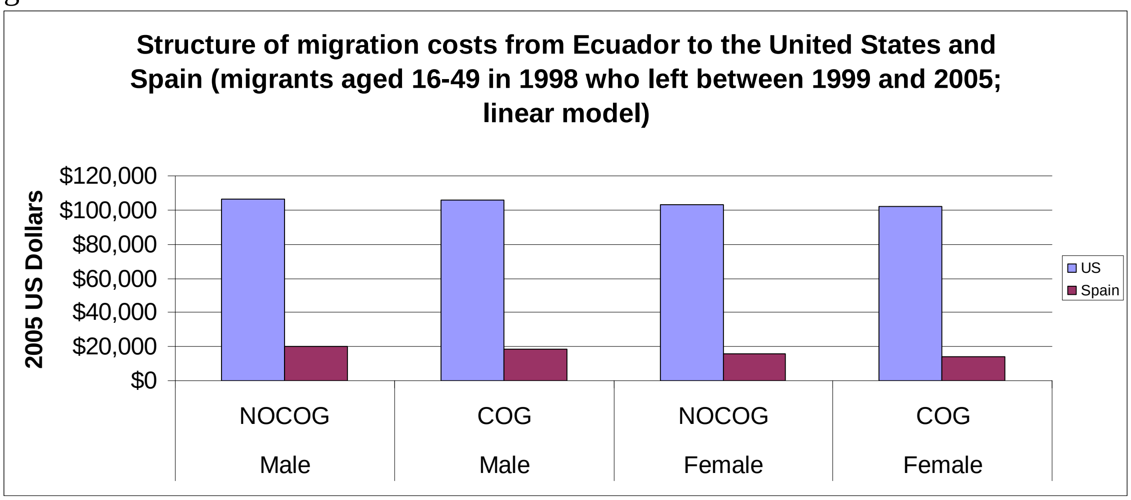


Figure 4:



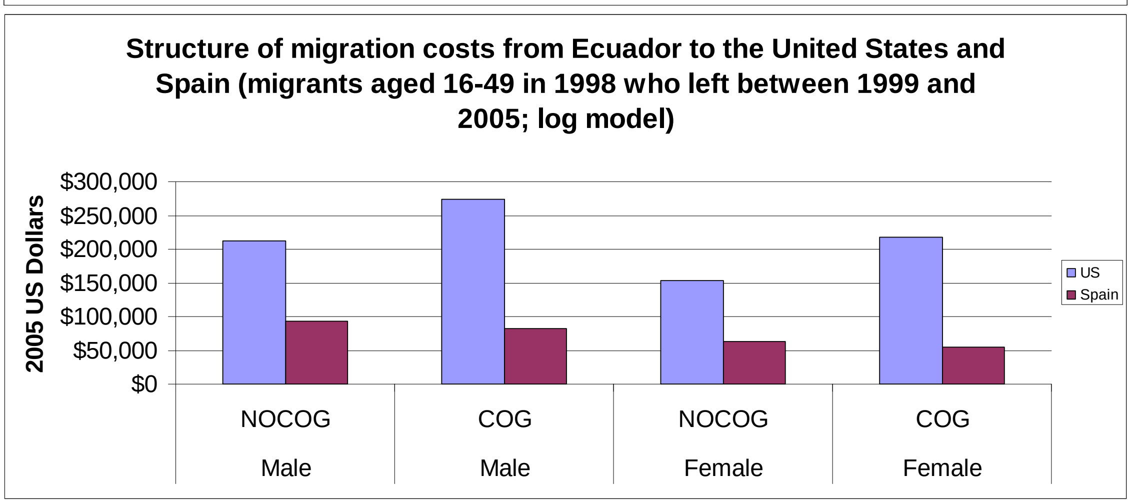


Figure 5:

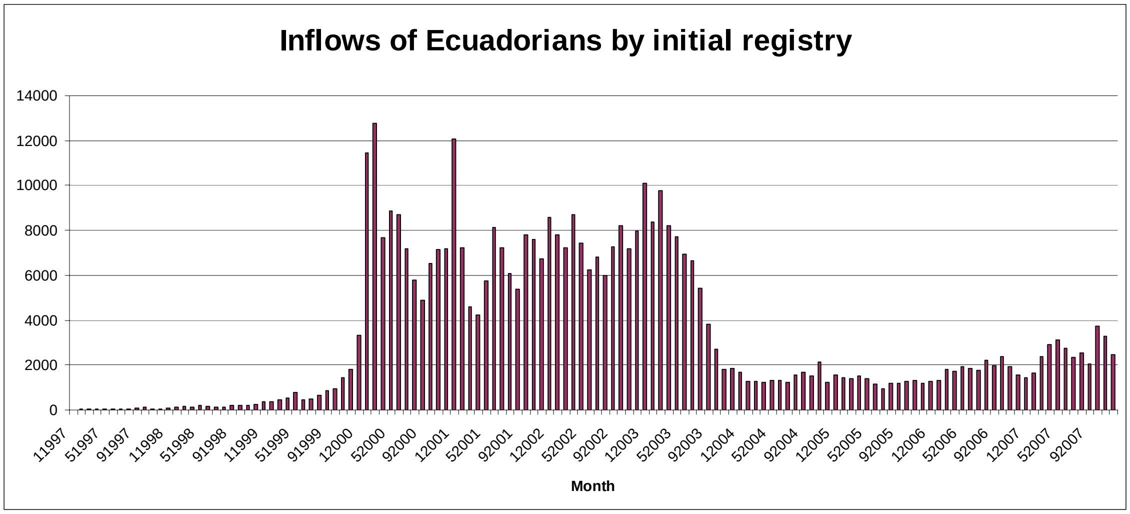


Figure 6: