Crossing the border: Self-selection, earnings and individual migration decisions

S. Bertoli [a,*], J. Fernández-Huertas Moraga [b], F. Ortega [c]

[a] CERDI, University of Auvergne, Bd. F. Mitterrand, 65, F-63000, Clermont-Ferrand, France
[b] FEDEA, Jorge Juan, 46, E-28001, Madrid, Spain
[c] Queens College CUNY, Department of Economics, 65-30 Kissena Boulevard, Flushing, NY 11367, USA

ARTICLE INFO

ABSTRACT

We analyze an international migration episode 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 of 1999. Specifically, we produce selection-corrected predictions of counterfactual individual earnings and use them to estimate a discrete-choice migration equation that allows for correlated errors across destinations and a rich structure of migration costs. We find that earnings significantly shape individual migration decisions, even in an episode in which Ecuadorians mostly chose Spain where earnings were lower than in the US, and they contribute to explaining the observed composition of migration flows. Moreover, our estimates show that changes in earnings at a particular destination have a larger effect on destination choice conditional on migration than on the scale of migration.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Why do people move across borders? And, in particular, what is the role of income differences in determining international migration flows? This is a simple yet challenging question, going back to Sjaastad (1962). Estimating the role of income in migration decisions requires predicting the earnings individuals can obtain at all alternative locations. But, of course, each individual is only observed in one single location.

A number of recent influential studies have made important contributions toward understanding the role of income in accounting for bilateral migration flows (Grogger and Hanson (2011), Belot and Hatton (2012), Ortega and Peri (2009), among several others). These studies typically use solely aggregate data, so that country-wide average income figures — specifically, GDP per capita — are used to proxy potential migrants' earnings at destination. This choice — which is severely constrained by data availability — implicitly rests on two assumptions, namely that destination countries do not differ as far as the transferability of migrants' skill is concerned, and that there are no individual-specific unobserved factors that simultaneously influence earnings and the decision to migrate. Still, none of these two assumptions is fully consistent with the findings of the empirical literature on

the assimilation of immigrants (see Chiswick (1978) and Borjas (1985)), and the theoretical insights on migrants' self-selection (see Roy (1951), Sjaastad (1962) or Borjas (1987)).

On the other hand, the internal migration literature (e.g. Dahl (2002); Bayer et al. (2009); Kennan and Walker (2011)) employs individual-level data. It is a well-established fact in the internal migration literature that failing to account for unobserved ability or any other factor that affects simultaneously the migration decision and expected earnings can represent a critical source of bias.

Our contribution is the estimation of an international migration model using individual-level earnings for all relevant locations collected from different sources. The model allows for unobserved individual-specific factors influencing both earnings and migration decisions, as in the Roy-Borjas model. In the estimation of the earnings equation, we control for self-selection using state-of-the-art techniques (Dahl 2002) from the internal migration literature. In the estimation of the migration decision, while controlling for a rich structure of migration costs, we relax the independence of irrelevant alternatives assumption,[1] so that migration decisions do not respond to changes in earnings differentials but differentially to changes in earnings in the various locations.

* Corresponding author.
E-mail addresses: simone.bertoli@u-clermont1.fr (S. Bertoli),
jfernandezhuertas@fedea.es (J. Fernández-Huertas Moraga), fortega@qc.cuny.edu
(F. Ortega).

[1] De Vreyer et al. (2010) also use individual-level data collected in seven capital cities in Western Africa; they control for self-selection into migration, but their estimated migration model keeps the independence of irrelevant alternatives assumption.

Table 1
Ecuadorian migration flows between 1999 and 2005.

Ecuadorian migrants born in 1949–1982			
	Destination country		Ratio
	US	Spain	Spain/US
Males	44,410	124,586	2.8
	(3035)	(7455)	(0.3)
Non-college	37,927	115,069	3.0
	(2895)	(7312)	(0.3)
College	6183	9517	1.5
	(1030)	(1790)	(0.4)
Females	37,914	121,352	3.2
	(2605)	(6558)	(0.3)
Non-college	29,705	102,935	3.5
	(2342)	(6131)	(0.4)
College	8209	18,417	2.2
	(1243)	(2699)	(0.5)
Total	82,024	245,938	3.0
	(3788)	(9148)	(0.2)
Non-college	67,632	218,004	3.2
	(3571)	(8909)	(0.2)
College	14,392	27,934	1.9
	(1599)	(3212)	(0.3)

Notes: Authors' elaboration on ACS 2007 and ENI 2007; sampling weights used to compute the figures; standard errors in parenthesis.

This approach is used to analyze a recent major migration episode, namely the wave of Ecuadorian migration that was triggered by the late 1990s economic crisis, when approximately 600,000 individuals left from a country with a population of 12.7 million over a few years (1999–2005). A unique feature of this migration episode is that most Ecuadorian migrants moved to just two destinations. Together the US and Spain received about 85% of all Ecuadorian migrants. We merge information on Ecuadorians contained in three comparable household surveys collected in Ecuador, the National Survey on Employment and Unemployment in the Rural and Urban Areas 2005 (ENEMDU 2005), the US, the American Community Survey 2007 (ACS 2007), and Spain, the National Immigrant Survey 2007 (ENI 2007).

This particular international emigration episode is also interesting for another reason. Namely, the number of Ecuadorians that migrated to Spain over our period of interest was roughly three times larger than the corresponding flow to the US (Table 1). This is puzzling given the large difference in per capita incomes between the two destinations[2] and the existence of pre-crisis Ecuadorian migration networks in the US but not in Spain (Jokisch and Pribilsky 2002). Thus this episode poses a challenging test for any income-maximizing theory of migration.

Our main result is that earnings differences were relevant determinants of the decision to migrate, even in an episode where most migrants preferred a lower income destination (Spain) over a higher income one (the US). The estimates also show that changes in earnings at a particular destination have a larger effect on destination choice conditional on migration than on the scale of migration. Econometrically, this is the main payoff of adopting the more general nested logit structure, which captures the correlation across destinations (conditional on migration) generated by individual heterogeneity in the unobserved propensity to migrate.

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. The application of the selection-control procedure proposed by Dahl (2002) allows us to derive counterfactual earnings which are

consistent even in the presence of unobserved individual heterogeneity affecting both wages and the migration decision. Reassuringly, in this particular case, the bias connected to the reliance on simple Mincerian regressions does not appear to be large. Still, the absence of an exogenous instrument induces us to acknowledge that some bias might remain even with the adoption of the approach proposed by Dahl (2002), which marks an improvement over standard techniques. Identification in the first stage (the wage equation) relies on the non-parametric characteristics of the Dahl procedure. In order to achieve identification in the second stage (the migration equation), we use broad occupational categories to predict earnings while omitting them from the nested logit estimation.[3]

The time-equivalent implicit migration costs that we recover from the model imply that the cost of moving to the US is several times larger than the corresponding cost of going to Spain. This difference could be related to the cultural and linguistic proximity between Ecuador and Spain, and to the relatively more generous Spanish welfare state. We provide suggestive evidence showing 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 August 2003) also contributed to shape the pattern of moving costs, and policy-induced migration costs (Beine et al., 2011) indeed constrained location choice. Networks, which did play a role in this migration episode (Bertoli, 2010), probably lowered the cost of moving to the US, but their effect is overshadowed by the influence of the other country-specific factors described above. The variability of migration costs across gender and educational groups is in line with the models put forward by Chiquiar and Hanson (2005) or McKenzie and Rapoport (2010).

This paper is related to several strands of research. It is most relevant in terms of the recent work on the determinants of international migration flows, such as Grogger and Hanson (2011), Belot and Hatton (2012), Ortega and Peri (2009), Mayda (2010), Pedersen et al. (2006), Clark et al. (2007) and Bertoli and Fernández-Huertas Moraga (2011). As noted earlier, all these studies rely on aggregate data.[4] In our use of individual earnings data for different countries, our work is in the vein of Clemens et al. (2008) who report wages of comparable workers with the same nationalities across different destination countries, and Batista (2008) who controls for unobserved heterogeneity when estimating individual-level counterfactual wages. In a unique, recent study, McKenzie et al. (2010) provide estimates of the gain from migration based on experimental evidence (a visa lottery). Their results suggest that it is important to control for selective migration, both in terms of observable and unobservable characteristics. None of these studies estimates a migration decision equation. Our paper is also related to other studies in the international migration literature that use micro data. Chiquiar and Hanson (2005), McKenzie and Rapoport (2010) or Fernández-Huertas Moraga (2011) study Mexico-US migration but are not concerned about the income-sensitivity of migration. Hanson and McIntosh (2010) also deal with Mexican emigration to the US and assess the factors behind long-run trends in the flows, establishing the relevance of labor supply shocks. Their work is extended to Latin American emigration in general in Hanson and McIntosh (2012). One of the channels through which labor supply shocks could be

---

[2] The 2006 GDP per capita in PPP US Dollars was 44,000 in the US, 29,000 in Spain and 7,000 in Ecuador (World Bank, 2008).

[3] We discuss below how the use of the nested logit takes care of the correlation induced by the reliance on occupations to predict counterfactual earnings while the potential endogeneity problem in the migration equation is not empirically relevant in this particular case.

[4] Yang (2006) represents an exception, as he uses individual-level data from a single data source to estimate the sensitivity of return decisions of Filipino migrants to the income shocks due to exchange rate fluctuations.