On the Determinants of Bilateral Migration Flows*

Jesus Crespo Cuaresma[†] Mathias Moser[‡] Anna Raggl[§]

Preliminary version. Please do not cite yet.

Abstract

We present a method aimed at estimating global bilateral migration flows and assessing their determinants. We employ that fact that available net migration figures are (nonlinear) aggregates of migration flows from and to all other countries of the world to construct a statistical model that links the determinants of (unobserved) migration flows to total net migration. Using specifications based on simple gravity models for migration, we find that bilateral migration can for a large part be explained by standard gravity model variables such as GDP difference, distance or bilateral population.

Keywords: Bilateral migration flows, gravity model, nonlinearly aggregated models.

*The authors acknowledge funding from the Austrian Science Fund (FWF): Z171-G11.

[†]Department of Economics, Vienna University of Economics and Business (WU); World Population Program, International Institute of Applied Systems Analysis (IIASA); Wittgenstein Centre for Demography and Global Human Capital (WIC) and Austrian Institute for Economic Research (WIFO). Address: Augasse 2-6, 1090 Vienna (Austria). Email: jcrespo@wu.ac.at.

[‡]Department of Economics, Vienna University of Economics and Business (WU). Address: Augasse 2-6, 1090 Vienna (Austria). Email: matmoser@wu.ac.at

[§]Research Institute for Human Capital and Development, Vienna University of Economics and Business (WU) and Wittgenstein Centre for Demography and Global Human Capital (WIC). Address: Augasse 2-6, 1090 Vienna (Austria). Email: araggl@wu.ac.at

1 Introduction

In 1990, there were approximately 150 million international migrants in the world, a figure that increased by more than 40% in the following two decades. Currently, about 214 million people worldwide live outside the country where they were born, a number that represents roughly 3.1% of total population (see United Nations (2011)).

The lack of availability of global databases for bilateral migration flows is a barrier to the understanding of the causes and consequences of international migration. While the OECD's International Migration Statistics (IMS) dataset (OECD, 2009) provides data on bilateral immigration flows, the information is limited to migration to a relatively small group of industrialized economies. Beine, Docquier, and Ozden (2011) presents a dataset of bilateral migration stocks by educational attainment for over 170 countries in 1990 and 2000, which researchers have used to construct migration flows as differences between stocks at these two points in time (see for example Beine, Docquier, and Ozden (2011)). The problems involved in using differences in migration stocks as a proxy of migration flows can be important and are often acknowledged in the empirical studies performing such an approximation. Mortality and return migration distort the quality of such a variable as a measurement of migration flows and thus the assessment of the dynamics of newcomers based on the difference in the stock of migrants can lead to flawed inference.

Gravity models have often been used to assess empirically the determinants of migration flows using the data mentioned above, see for example Grigg (1977); Vanderkamp (1977); Clark, Hatton, and Williamson (2007); Karemera, Oguledo, and Davis (2000). Data availability thus tends to limit these empirical studies on the determinants of bilateral migration to cases where the recipient country is an advanced OECD economy.

In this study, we propose a new method that allows us to estimate bilateral migration flows and study their empirical determinants using net migration data, which are available for practically all countries in the world. In particular, we assume that bilateral migration flows can be described by a simple gravity model and construct a specification based on net migration, which can therefore be thought of as a nonlinear aggregation of these unobserved bilateral variables. Such a modelling strategy allows us to estimate the effects of the various determinants of bilateral migration and construct estimates of bilateral migration flows as the corresponding fitted values. Our approach presents a natural framework to obtain projections of bilateral migration flows that can be used to improve existing population projection exercises.

The paper is organized as follows. In section 2, we present the statistical modelling framework and describe the estimation of the underlying parameters. Section 3 carries out a simulation study to evaluate the method. Section 4 conducts the empirical analysis and obtains estimates of bilateral migration flows among 172 countries for the year XXX. Section 4 concludes.

2 Modelling nonlinearly aggregated bilateral migration flows

2.1 The underlying specification

Assume that (log) bilateral migration flows are represented by the model

$$m_{ij} = \log M_{ij} = X_{ij}\beta + u_{ij}, \tag{1}$$

where M_{ij} denotes migration from country j to country i, X_{ij} is a $1 \times k$ vector of determinants of bilateral migration, β is a $k \times 1$ vector of parameters to be estimated and u_{ij} is an error term assumed independent, identically distributed and homoskedastic with variance σ^2. Bilateral flows are not observed, but data for n countries exist on net migration (N_i), which is given the difference of migration flows to country i from all other countries and migration out of country i to all other countries,

$$N_i = M_{i*} - M_{*i} = \sum_{j \neq i} M_{ij} - \sum_{j \neq i} M_{ji} = \sum_{j \neq i} \exp m_{ij} - \sum_{j \neq i} \exp m_{ji}. \tag{2}$$

The model for our observed data can thus be written as

$$\mathbf{N} = \mathbf{S} \exp(\mathbf{m}) = \mathbf{S} \exp(\mathbf{X}\beta + \mathbf{u}), \tag{3}$$

where $\mathbf{N}$ is an n-dimensional column vector of net migration observations, $\mathbf{X}$ is an $(n-1)^2 \times k$ matrix of observations on the bilateral explanatory variables, $\mathbf{S}$ is an $n \times (n-1)^2$ matrix which selects the corresponding bilateral migration flows, aggregates them for each country and creates the net migration figures and $\exp(\mathbf{v})$ denote the element-by-element exponent of vector $\mathbf{v}$. Assuming that $\mathbf{m}$ is ordered by recipient country, a typical row k of $\mathbf{S}$ has elements equal to 1 in columns $(n-1)(k-1)+1$ to $(n-1)(k-1)+(n-1)$ and -1 in columns $k+k(n-1), k+(k+1)(n-1), \ldots$ for elements in positions beyond $k(n-1)$ and in columns $k-1+k(n-1), \ldots k-1+(k+1)(n-1)$ for elements prior to $(k-1)(n-1)$. This implies that the matrix $\mathbf{S}$ is given by $\mathbf{S} = (\mathbf{I}_n \otimes \iota_{n-1}) - \mathbf{B}$, where $\mathbf{B}$ is an $n(n-1)^2$ matrix composed by $n-1$ column blocks of matrices whose structure corresponds to augmenting the matrix $-\mathbf{I}_{n-1}$ by one row of zeros in position b for each block $b = 1, \ldots, n-1$.

While the model for the bilateral migration flows is linear in parameters, the aggregation of the flows which yields the net migration flows implies a nonlinear link between $\mathbf{N}$ and β. Therefore, we cannot estimate our model with least squares and rely on nonlinear maximum likelihood methods to estimate β.

2.2 Estimating the model

A simple approach to the estimation of model (3) implies ignoring the nonlinearity in the error term and estimating β based on a specification where the disturbance is defined at the level of the aggregated variable (N_i) instead of at the bilateral level,

$$\mathbf{N} = \mathbf{S} \exp(\mathbf{X}\beta) + \eta, \tag{4}$$

which allows to estimate β using nonlinear least squares or pseudo maximum likelihood methods. Assuming independence, normality and homoskedasticity for the disturbance term, the likelihood of the model can be written as

$$L(\beta, \sigma_\eta | \boldsymbol{y}) = \prod_{i=1}^{n} f\left(y_i | \beta, \sigma_\eta\right), \tag{5}$$

with the corresponding log-likelihood function

$$\ell(\boldsymbol{\theta} | \boldsymbol{y}) = \sum_{i=1}^{n} \ln f\left(y_i | \boldsymbol{\theta}\right). \tag{6}$$

Assuming normality of the errors (they are not!), we have

$$\begin{aligned}
\ell(\boldsymbol{\theta} | \boldsymbol{y}) &= \sum_{i=1}^{n} \ln \left[\frac{1}{\sigma\sqrt{2\pi}} \exp \frac{(u_i + v_i)^2}{2\sigma^2} \right] & (7)\\
&= -n \ln \sigma - n \ln\left(\sqrt{2\pi}\right) + \frac{\sum_{i=1}^{n} (u_i + v_i)^2}{2\sigma^2} & (8)\\
&= -n \ln \sigma - n \ln\left(\sqrt{2\pi}\right) + \frac{\sum_{i=1}^{n} \left(N_i - \sum_{j \neq i} X_{ij}\beta + \sum_{j \neq} X_{ji}\beta\right)^2}{2\sigma^2} & (9)\\
&= -n \ln \sigma - n \ln\left(\sqrt{2\pi}\right) + \frac{(\boldsymbol{u} + \boldsymbol{v})'(\boldsymbol{u} + \boldsymbol{v})}{2\sigma^2} & (10)
\end{aligned}$$

where

$$(\boldsymbol{u} + \boldsymbol{v}) = \left[\boldsymbol{N} - (\boldsymbol{I_n} \otimes \boldsymbol{\tau_n}) \exp\left(\boldsymbol{X\beta}\right) + (\boldsymbol{\tau_n} \otimes \boldsymbol{I_n}) \exp\left(\boldsymbol{X\beta}\right)\right] \tag{11}$$

3 Data

Data on the language, common borders, colonial history and distance of the countries to each other are obtained from the CEPII Gravity Dataset (Head, Mayer, and Ries, 2010). Net migration rates as well as GDP, unemployment rates, inflation and population figures are taken from the World Development Indicators. Data on age group specific educational attainment is taken from the IIASA-VID dataset[1] (Lutz, K.C., and Sanderson, 2007).

4 Simulation Results

In a first step we evaluate the method by using simulated data. In a simplified setting we assume that the data is generated by a process like

$$m = \boldsymbol{\alpha} + .1\boldsymbol{X_1} + 0.5\boldsymbol{X_2} - 0.5\boldsymbol{X_3} + \boldsymbol{u} \tag{12}$$

where $X_{1,2,3}$ are exogeneous bilateral matrices for 100 simulated countries which are drawn from a standard normal distribution. The noise term is calibrated to represent different

[1] The version available online is less recent than the one used for this study.

levels of R^2, ranging from 0.95 to 0.7. For these levels we aggregate the bilateral flow $\mathbf{m}$ according to equation 1 to receive simulated net migration flows $\mathbf{N}$.

In order to assess the quality of our method we reestimate the bilateral migration flows by using only the exogeneous variables and the net migration flows on the LHS of the equation. This is done by using standard maximum likelihood estimates. We repeat this exercise 1000 times for each noise level to ensure stability of our results. For every coefficient table 1 presents means and RMSE of the estimated coefficients.

Table 1 – Simulation results for different levels of noise

	$R =$	0.95	0.90	0.85	0.80	0.75	0.70
$\beta_0(1.0)$	RMSE	0.082	0.121	0.156	0.215	8.922	12.748
	Mean	1.01	1.02	1.02	1.04	0.65	0.11
$\beta_1(.1)$	RMSE	0.017	0.027	0.036	0.041	0.083	0.970
	Mean	0.10	0.10	0.10	0.10	0.10	0.06
$\beta_2(-.5)$	RMSE	0.027	0.039	0.050	0.067	0.863	1.970
	Mean	-0.50	-0.50	-0.50	-0.51	-0.55	-0.63
$\beta_3(.5)$	RMSE	0.026	0.039	0.050	0.066	1.659	2.180
	Mean	0.50	0.50	0.51	0.51	0.57	0.64

As can be seen from the results the method works well for noise levels which correspond to R^2 of about 0.8. Since the model is estimated in differences (equation 1), the intercept is only weakly identified. Therefore the deviations of $beta_0$ tend to be larger.

5 Empirical Results

[[TO BE WRITTEN]]

References

Beine, M., F. Docquier, and C. Ozden (2011): "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

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.

Grigg, D. (1977): "E. G. Ravenstein and the 'laws of migration'," *Journal of Historical Geography*, 3(1), 41 – 54.

Head, K., T. Mayer, and J. Ries (2010): "The Erosion of Colonial Trade Linkages after Independence," *Journal of International Economics*, 81(1), 1–14.

Karemera, D., V. I. Oguledo, and B. Davis (2000): "A gravity model analysis of international migration to North America," *Applied Economics*, 32(13), 1745–1755.

Lutz, W., S. K.C., and W. Sanderson (2007): "Reconstruction of Population by Age, Sex and Level of Educational Attainment of 120 Countries for 1970-2000," *Vienna Yearbook of Population Research*, 2007, 193–235.

OECD (2009): "International Migration Statistics," Discussion paper, ESDS International, University of Manchester.

United Nations (2011): "International Migration Report 2009: A Global Assessment," ST/ESA/SER.A/316, Department of Economic and Social Affairs, Population Division.

Vanderkamp, J. (1977): "The Gravity Model and Migration Behaviour: An Economic Interpretation," *Journal of Economic Studies*, 4(2), 89–102.