DOI: 10.20472/EFC.2017.007.018

TATIANA POLONYANKINA

University of Economics, Prague, Czech Republic

DOES THE GRAVITY MODEL WORK FOR THE MODELLING OF MIGRATION BETWEEN EUROPEAN COUNTRIES FROM 2011 TO 2014?

Abstract:

The gravity model is an interesting adaptation of Newton's law of gravitation, in which the effect of gravity is used to describe the spatial interactions between economic units. The force of interaction is supposed to be positively influenced by the size of the units (the push factor) and negatively by the distance between them (the pull factor). The model is used to estimate the dependence of migration on the GDP, as well as the distance between European countries. Based on the gravity model, the GWP of both (source and host) countries, is expected to be a push factor and the distance is expected to be a pull factor. However, in economic theory, the impact of the GDP of a source country is expected to be negative, the opposite to the gravity model. The goal of the paper is to test which of the two is valid for eight European countries from 2011 to 2014.

Keywords:

gravity model, spatial dependence, migration, model selection, random effects, panel data

JEL Classification: F22, C23, C52

Introduction

"Spatial interaction" is a term used to describe any movement through space that is caused or forced by human interaction or other human activity. A common example is any kind of migration (from workers traveling every day for work to immigrants going to live in a host county). Other examples are commodity flow, international trade or knowledge exchange.

The gravity model is used to explain the spatial interaction between countries, since it includes geographical aspects. The model derives its name and the basic idea from physics. The main idea of the gravitation interaction between spatial bodies is that there are two main influences: the distance between the bodies and their scale impact. The scale can be measured in absolute values (size) or in relative values, depending on the field of the applied model (e.g. GDP as a measure for the economic field). Distance can also be defined as an absolute location – the distance between two places in kilometres/miles; or as a relative location – locations with the same absolute distance but with different access to shopping centres, schools, museums, access to job opportunities or information, and even access to water. The different fields of the model applications are described in the section named The Theory of Gravity Model.

The aim of this article is to test whether the assumptions of the gravity model are fully met in the case of migration to European countries, or whether the economic explanation of push factors is stronger, thereby making the impact of the GDP of the source country negative. The testing is done for eight countries of the European Union between 2011 and 2014 and is described in the section called Empirical Analysis.

The Theory of the Gravity Model

The gravity model supposes that size (in relative or absolute terms) has a positive effect on spatial interaction. Its modification for migration analysis assumes that the GDP (as a relative size measurement) of both the host and source country has a positive impact on migration. Economic theory and the gravity model diverge when considering the impact of the GDP of a source country on migration. The gravity model expects a positive effect of the GDP of a source country on migration, however, from an economic point of view, a negative impact is expected. Economic theory focusing on the determinants of migration says that migration is often motivated by expectations of a better economic situation of the individual in the host country. It is generally assumed that economically motivated individuals will move from economically weaker countries with a lower GDP to countries with a higher GDP. This is empirically examined in the study by Mayda (2003).

Constantin (2004) defines three modifications of the gravity model for three different application fields: A) the trade flows where the gravity model is applied as the Rielly law. The law says that a retail client is attracted to a shopping centre proportionally to the shopping centre size and inversely to the square of the distance from the shopping centre. B) the gravity model modification to analyse a commodity flow, where the flow depends directly on the demand for the commodity and indirectly on transaction costs. C) the gravity model for a migration analysis when migration is directly dependent on population or economic strength of the centre and indirectly on the distance from the centre.

The use of gravity models for spatial interaction analysis is summarised in Constantin (2004), where he talks about the market area boundary analysis, gravity models for commodity flow analysis and gravity models for studying the migration flows.

The classical gravity model was first used to explain trade flows by Tinbergen (1962). Later on, the model was widely used for international trade analysis (Anderson (1979)). The other macroeconomic field in which the gravity model is widely used is in the analysis of foreign direct investment (FDI) (Egger (2004)). Braconier (2005) explores how wage costs for high skilled and less skilled labour in host countries are affected by FDI using the gravity model. For migration flows, the gravity model was first used by Ravenstein (1889) for the UK. Metulini (2013) uses the spatial gravity model for international trade working with a panel of OECD countries and showing that the gravity model provides better results than the classical ordinary least squares regression (OLS) in which estimations are not unbiased and suffer from endogeneity.

Next, we define the basic gravity model using distance and scale measured in original values. All of the possible fields of the model applications were described earlier; however, from this section on, we only focus on the model modification for migration.

The gravity model assumes a spatial correlation, so it is reasonable to use when the researcher presumes that there can be dependencies between cells (also called observations or regions) due to its geographic location. The requirement of geographic information about the cell location in the dataset used is necessary. This information is captured in an origin – destination regions distance matrix.

The basic gravity model is defined by equation (1) where the interaction T_{ij} between two regions i and j is a function of P_i, the scale impact of region i, P_j, the scale impact of region j and d_{ij}, the distance between two regions, the constant G, and the parameters a, b.

$$T_{ij} = G\frac{{P_i}^a {P_j}^b}{{d_{ij}}^c} \qquad (1)$$

To estimate this equation, we need to log-linearize function (1) and add an error term. Doing so, we end up with:

$$logT_{ij} = logG + alogP_i + blogP_j - clogd_{ij} + \varepsilon_{ij} \qquad (2)$$

To make it easier to work with the equation, we do a substitution in equation (2) and get:

$$y_{ij} = \alpha + \beta_d x_{di} + \beta_o x_{oi} + \gamma d_{ij} + \varepsilon_{ij} \qquad (3)$$

It is assumed that the model has a total number of n regions. The interaction between regions is described by an $n\,x\,n$ square matrix of interregional flow from each of the n origin regions to each of the n destination regions. We can restructure this $n\,x\,n$

matrix into a $n^2 x\, 1$ vector y by converting the columns of the matrix into a variable vector. After the transformation, the final equation is derived:

$$y = \alpha J + X_d\beta_d + X_o\beta_o + D\gamma + \varepsilon, \quad (4)$$

where:

- y is an $n^2 x\, 1$ vector of logarithms of dependent variables which define interactions between cells,
- J is an $n^2 x\, 1$ intercept vector,
- X_d is an $n^2 x\, k$ vector of logarithms of destination region scale characteristics,
- X_o is an $n^2 x\, k$ vector of logarithms of origin region scale characteristics,
- D is an $n^2 x\, 1$ vector of logarithms of origin – destination region distances,
- $\varepsilon \sim N(0, \sigma^2 I_n)$,
- $\alpha, \beta_d, \beta_o, \gamma$ are parameters to be estimated.

Gravity Model Transformation for Migration Analysis

Variables in the basic gravity model, as defined in (1), are replaced by variables that are generally used in model modifications to analyse migration:

T_{ij} for interactions between *i* and *j* is replaced by migration from state *i* to state *j*, P_i for the scale impact of *i* is replaced by the GDP of state *i*, P_j for the scale impact of *j* is replaced by the GDP of state *j*, d_{ij} remains as the distance between states *i* and *j* (measured as the distance from the centre), and the constant *G* and parameters a, b, c remain the same as well:

$$migration_{ij} = G\,\frac{GDP_i{}^a GDP_j{}^b}{d_{ij}{}^c} \quad (5)$$

The GDP is measured in per capita, so to include only economic power and not the population size of a country. Again, log linearization and model transformation is done and a time index is added to derive the final regression equation for panel data where $n = 8, k = 1$ and $t = 2011 - 2014$:

$$y_t = \alpha J + X_{it}\beta_o + X_{jt}\beta_d + D\gamma + \varepsilon_t, \quad (6)$$

where:

- y is an $n^2 x\, 1$ vector of logarithms of dependent variables which defined migration between countries in year t,
- J is an $n^2 x\, 1$ intercept vector,
- X_{it} is an $n^2 x\, 1$ vector of logarithms of GDP per capita of source countries in year t,
- X_{jt} is an $n^2 x\, 1$ vector of logarithms of GDP per capita of host countries in year t,
- D is an $n^2 x\, 1$ vector of logarithms of distances between host and source countries (in km),
- $\varepsilon_t \sim N(0, \sigma^2 I_n)$,
- $\alpha, \beta_d, \beta_o, \gamma$ are parameters to be estimated.

Econometric model

The model (6) is doing to be used to estimate the dependence of migration between European countries. To get unbiased and robust estimates of the model parameters we need to use techniques to fit econometric model verification.

Since the model is working with macroeconomic time series a test for stationarity test must be done all variables. In case of non-stationarity of used variables, a problem of seeming regression can appear when uncorrelated variables are defined as correlated in the regression output (Arlt (2002)).

Let's assume we have a time series process s_t type AR (1), which can be written as:

$$s_t = \rho s_{t-1} + u_t, u_t \sim N(0, \sigma^2) \quad (7)$$

Unit root tests are used for stationarity testing. Tests have a null hypothesis of unit root versus an alternative with a single stationary value.

$H_o: \rho = 1; s_t \sim\ I(1) =>$ non-stationarity

$H_1: |\rho| \leq 1; s_t \sim\ I(0) =>$ stationarity

One of the possibilities to test this is the Harris-Tzavalis unit-root test. It's designed to be applied to data sets which are relatively short in time. Other words it assumes that the number of panels tends to infinity while the number of time periods is fixed.

When the testing process is concluded as non-stationary, a transformation before using it is regression need to be done.

$$\Delta s_t = s_t - s_{t-1}, t = 2, \dots T \quad (8)$$

One way how to deal with the non-stationarity problem is first difference transformation. The transformation is described with equation (8) and is used in this article.

The analysis using panel data allows to allows to control for variables which cannot be observe or measure like cultural factors or difference in business practices across companies; or variables that change over time but not across entities (i.e. national policies). Two main techniques are used to analyse panel data: Fixed effects and Random effects.

Fixed-effects are used when researcher is only interested in analysing the impact of variables that vary over time. Using random effects is possible to include time invariant variables (i.e. distance). The random effects can be defined as:

$$y_{it} = X_{it}\beta + \alpha + u_{it} + \varepsilon_{it},\ \ u_{it} \sim N(0, \sigma^2), e_{it} \sim N(0, \sigma^2) \quad (9)$$

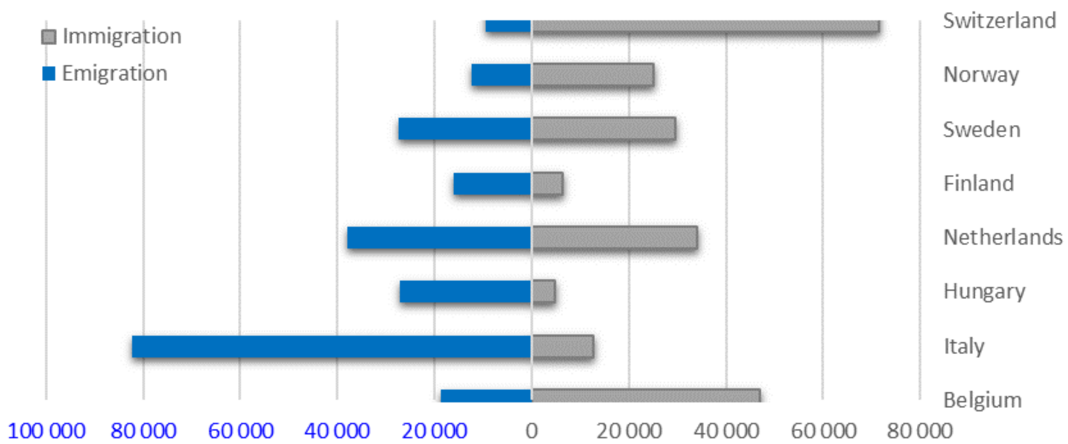
Random effects assume that the entity's error term u_{it} is not correlated with the predictors which allows for time-invariant variables to play a role as explanatory variables.

Empirical Analysis

We use the model described in (6) to estimate the dependence of migration, within 8 European countries, on their GDP and the distances to each other. The migration data source is the Eurostat web page. The GDP per capita source is the Czech Statistical Office database. Countries were chosen based on Eurostat data availability since the model requires information about migration flows for all countries. Regression analysis is based on panel data for the years 2011-2014 and countries: Belgium, Italy, Hungary, the Netherlands, Finland, Sweden, Norway and Switzerland.

Figure 1 shows the sum of migration flows between European countries from 2011 to 2014 where a source country is defined as the one where an individual was born. The country with one of the lowest immigration rates in comparison to the other seven countries is Hungary. The countries that show a high increase of immigrants from other countries considered are Belgium, Switzerland and the Netherlands. The country with the highest emigration is Italy. This outflow of citizens is likely related to the high unemployment rate Italy experienced during the economic crisis, which forced young people especially to go abroad.

Figure 1: Migration between European countries from 2011 to 2014



Source: Eurostat public database 2/2017.

Before estimating the regression, a test for stationarity was done for all variables, since the paper deals with macroeconomic time series. The Harris-Tzavalis unit-root test assumes that the number of panels tends to infinity while the number of time periods is fixed. Under this assumption, we can conclude that the test is suitable for panel data with a large number of panels and a small number of time periods. The test concludes that both GDP variables X_{it} and X_{jt} do not meet the stationarity assumption and a first difference transformation of these variables is done to avoid a seeming regression problem.

Panel data analysis is done using Stata with a random effects transformation. Since the model includes only three explanatory variables, it is reasonable to believe that differences across cells (countries) have some influence on migration (the dependent variable), which is why we used a random effects transformation. The second possibility – a fixed effects transformation, was not used since we would lose one of our variables, namely distance (D), since it does not vary across time, and thus this variable would be absorbed by the intercept in the fixed effects model.

The regression also includes a robust option for estimating the standard errors using the Huber-White sandwich estimators to deal with potential heteroscedasticity.

Final parameter estimations are displayed in Table 1. The overall R-squared is 0.2145. The results show that distance has a negative and statistically significant impact on migration, which was expected from both the economic theory and the gravity model. The GDP of a destination has a positive impact, significant only at the 10% significance level. This result is also in line with both the economic theory and the gravity model.

The impact of the GDP of a source country on migration is negative, but not statistically significant. The gravity model assumptions do not fit the migration analysis results since the gravity model expects a positive and statistical significant impact. The results tend to support the economic theory regarding the negative impact, but the regression results are not strong enough to get a statistically significant result. The selection of European countries used in the analysis includes stable and developed economies where economic migration push factors are not as visible as in the case of migration from economically weaker countries outside of Europe.

Table 1: Regression results

explanatory variables	coef.	Robust std. Err.	Z	p-value
constant	12.238	2.055	5.95	0.000
1. diff log GDP source country	-0.038	0.061	-0.63	0.531
1. diff log GDP destination country	1.575	0.139	1.80	0.071
log distance	-0.864	0.029	-2.95	0.003

Random-effects GLS regression Number of obs. = 168
Group variable: cell Number of groups = 56

Source: Own calculation is Stata.

Conclusion

The paper analysed migration flows between eight European countries between 2011 and 2014 using the gravity model. The model is based on Newton's law of gravity, which says that the interaction between two bodies depends positively on their size and negatively on their distance. The basic gravity model is adapted to migration analysis where the economic strength defined by GDP per capita is used. The panel data regression with random effects and robust standard errors was used.

The goal of the paper was to see whether the assumptions of the gravity model work in its application to panel data analysis of migration.

The estimations of the parameters for the GDP of the destination country and distance meet the requirements of both the gravity model and also the economic theory. Specifically, a positive and statistically significant effect of the GDP of a host country was found, as was a negative and statistically significant effect of the distance between countries.

The gravity model assumptions were met for the impact of the distance and the GDP of the host country on migration between European countries, but were not valid for the impact of the GDP of the source country.
The gravity model assumes a positive and statistically significant effect of the GDP of the source country. Economic theory assumes the opposite effect, i.e. emigration is higher in countries with lower GDP where individuals expect that immigration to economically stronger countries will improve their economic situation. The negative estimated parameter for the GDP of a source country seems to support the economic theory, however the parameter is not statistically significant. European developed countries were used in the analysis, therefore migration pull factors may not be as obvious as emigration from economically weaker countries.

Acknowledgements

The work on this paper was supported by the grant IGS F4/73/2016 of the University of Economics, Prague.

Reference

ANDERSON, J. A. (1979). A Theoretical Foundation for the Gravity Model, *The American economic Review*, Vol. 69, p. 106-116.

ARLT, J., ARLTOVÁ M. and RUBLÍKOVÁ E. (2002). *Analýza ekonomických časových řad s příklady.* 1.vyd. Praha: Vysoká škola ekonomická Praha, 14 s. ISBN 80-245-0307-7.

BORJAS G. J. (2010). *Labour Economics*, Boston et al.: McGraw-Hill, ISBN: 9780073511368.

BORJAS G. J. (2005). The Labour Market Impact of High Skill Immigration, *American Economic Review*, Vol. 95, No. 2, p. 56-60.

BRACONIER, H., NORBÄCK, P. J., and URBAN, D. (2005). Multinational enterprises and wage costs: Vertical FDI revisited. *Journal of International Economics*, Vol. 67, No.2, p. 446-470.

CONSTANTIN, D. L. (2004). The use of Gravity Models for spatial interaction analysis. *Economy Informatics*, Vol.1, No.4, p. 116-118.

EGGER, P., and PFAFFERMAYR, M. (2004). Foreign direct investment and European integration in the 1990s. *The World Economy*, Vol. 27, No. 1, p. 99-110.

HAYNES, KINGSLEY E. and A. STEWART. FOTHERINGHAM. *Gravity and spatial interaction models.* Beverly Hills: Sage Publications, c1984. ISBN 0803923260.

LESAGE J., and PACE R. K. (2009). *Introduction to spatial econometrics*, CRC press, FL 33487-2742.

LESAGE J. (2005). Lecture 7: Spatial Econometric Modelling of Origin-Destination flows. Lecture notes.

MAYDA, A. (2007). International migration: A panel data of the determinants of bilateral flows. *Quarterly Journal of Economics.* Vol. 116, No. 2, p. 549-599.

METULINI R. (2013). Spatial gravity models for international trade: a panel analysis among OECD countries, *European Regional Science Association conference paper*, ersa13p522.

RAVENSTEIN, E. G. (1889). The laws of migration. *Journal of the royal statistical society*, Vol. 52, No.2, p. 241-305.

TINBERGEN, J. (1962). Shaping the world economy; suggestions for an international economic policy. *Books (Jan Tinbergen).*

TORRES-REYNA, O. (2007). Panel data Analysis Fixed and Random Effects using Stata. http://dss.princeton.edu/training/