Ramos, Raul

Article

Gravity models: A tool for migration analysis

IZA World of Labor

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Ramos, Raul (2016) : Gravity models: A tool for migration analysis, IZA World of Labor, ISSN 2054-9571, Institute for the Study of Labor (IZA), Bonn, https://doi.org/10.15185/izawol.239

This Version is available at:
https://hdl.handle.net/10419/148464

RAUL RAMOS
University of Barcelona, Spain, and IZA, Germany

I Z A
World of Labor
Evidence-based policy making

Gravity models: A tool for migration analysis

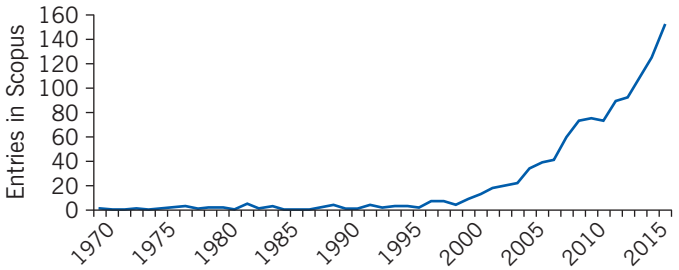
Availability of bilateral data on migratory flows has renewed interest in using gravity models to identify migration determinants

Keywords: gravity models, international migration, migration determinants, migration policies

ELEVATOR PITCH

Gravity models have long been popular for analyzing economic phenomena related to the movement of goods and services, capital, or even people; however, data limitations regarding migration flows have hindered their use in this context. With access to improved bilateral (country to country) data, researchers can now use gravity models to better assess the impacts of migration policy, for instance, the effects of visa restriction policies on migration flows. The specification, estimation, and interpretation of gravity models are illustrated in different contexts and limitations of current practices are described to enable policymakers to make better informed decisions.

Gravity models have become widely used by academics and policy advisors



Note: Shows the number of entries in Scopus when using the search term "gravity model" in the "economics" field.

Source: Author's own elaboration from Scopus data. Online at: http://www.scopus.com. Retrieved January 8, 2016.

I Z A
World of Labor

KEY FINDINGS

Pros

- Gravity models provide an intuitive framework to understand the determinants of flows between countries, in particular: trade, migration, or capital.
- Gravity models can easily be derived from theoretical models such as random utility maximization models.
- There are multiple methods to account for analytical challenges associated with gravity models such as the use of instrumental variables or fixed effects.
- Empirical models can easily be augmented to consider different additional controls and policy variables.

Cons

- The estimation of gravity models requires country-pairs detailed data (i.e. data regarding two specific countries experiencing direct flows between them), which is not always easy to obtain.
- Gravity models encounter difficulty when using data sets that include negative or zero values; some solutions are being investigated, but the challenge remains.
- The interpretation of gravity model results from a policy perspective is not always straightforward due to questions regarding data completeness and other influencing factors.

AUTHOR'S MAIN MESSAGE

Gravity models assume that flows between two countries are directly proportional to their size (population or GDP) and are inversely proportional to the physical distance between them (similar to Newton's gravitational law). Due to the recent availability of bilateral (i.e. two-way, country to country) migration data, gravity models have become more frequently used in the context of migratory flows. This allows for a better understanding of migration determinants when assessing policy impacts. Further improvement in the application of new data sets will enhance the usefulness of gravity models in a migration policy context.

MOTIVATION

The simplest versions of gravity models relate bilateral migration to the relative size of the origin and destination countries and the distance between them; however, there are additional factors that can affect migration flows. For this reason, gravity models are enlarged with variables related to different migration pull and push factors; for instance, better economic opportunities in the destination country (i.e. prospects for higher wages or lower unemployment rates), safer conditions, and higher political freedom, among others.

Gravity models have been used to understand the role of exogenous factors such as distance or linguistic proximity, while also being used to assess policy impacts such as visa restrictions. In fact, the use of gravity models has been growing extensively during the last decades, although there are still some limitations in terms of data availability and other technical issues.

DISCUSSION OF PROS AND CONS

Micro-foundations of the gravity model of migration

The theoretical basis for gravity models of migration is generally represented by a random utility maximization (RUM) model (see [1], [2], and [3], among others). RUM models describe the utility that an individual receives from living in a particular country compared to the expected utility received if moving to alternative destinations. The comparison involves both expected benefits (i.e. factors increasing the attractiveness of the destination such as higher expected earnings) and costs of migrating from origin to destination (such as distance or unfavorable migration policies).

The RUM model also includes a component that captures the unobserved factors of the individual utility associated with each choice. The researchers' assumptions regarding the statistical uncertainty around this component determine the expected probability that an individual will maximize his/her utility by opting for a particular destination. For instance, a logit-normal distribution can be adopted in such a way that the expected gross migration flows from one country to any other depend on the characteristics of the origin country, the attractiveness of the destination, and the accessibility of the destination country for potential migrants, which are criteria that clearly resemble a gravity model [1]. One relevant assumption of the RUM model is that the attractiveness of a destination is not supposed to be affected by migration. For instance, if one particular destination is attractive due to its low levels of unemployment when compared to a particular origin, massive inflows of immigrants could increase unemployment in the destination while at the same time decreasing it in the origin country. Gravity models do not capture these second-round effects, which is an important point to consider in order to appropriately interpret the model's results.

In any case, RUM models provide an appropriate theoretical justification of the intuition behind gravity models. The use of RUM models makes clear what the assumptions made by researchers are, and how these assumptions yield different specifications of empirical gravity models.

Gravity models of migration: Available data sets

Bilateral migration flows and stocks

UN Global Migration database
Includes information on the evolution of international migrants by country of birth and citizenship based on different sources such as population censuses, population registers, nationally representative surveys and other official statistical sources. Flow estimates are presented for 1990, 2000 and 2010 for more than 200 countries.

Online at: http://www.un.org/en/development/desa/population/migration/data/

World Bank Global Bilateral Migration database
Census and population register records are combined to construct matrices of bilateral migrant stocks for 1960, 1970, 1980, 1990, and 2000. Foreign-born definition of migrants is used. Bilateral migration matrixes for a reduced set of countries are also provided for 2010 and 2013.

Online at: http://go.worldbank.org/092X1CHHD0

OECD DIOC-E
Contains data on 89 countries of residence and covers all individuals aged 15 and over living in these countries. For most countries the place of birth is used to identify migrants, although in some cases it was necessary to rely on criteria based on nationality. The database identifies 232 countries of origin. It only contains information of migrant stocks, but provides detailed information on the educational level of immigrants, although it is not possible to control for the geographic location where the education or training was received.

Online at: http://www.oecd.org/migration/databaseonimmigrantsinoecdandnon-oecd-countriesdioc-e.htm

Geographical variables and additional controls and policy variables

CEPII GeoDist
Provides data on several geographical and other variables that can be used to estimate gravity models. Different measures of bilateral distances are available for 225 countries. It incorporates country-specific geographical variables, including capital cities coordinates, languages spoken in the country, a variable indicating whether the country is landlocked, and colonial links. Different measures of bilateral distances are also available for most country pairs across the world. CEPII's Gravity data set adds some additional time-varying variables for the period 1948–2006 to the GeoDist data set. In particular, data for GDP, population, and other institutional variables such as regional trade agreements and currency unions are also provided.

Online at: http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=6

UN World Population Policies database
Provides information about the evolution of government views and policies regarding different demographic dimensions, including internal and international migration.

Online at: http://esa.un.org/PopPolicy/about_database.aspx

Measurement issues (I): Country of birth, citizenship, or country of residence?

Due to some specific limitations in the data sets available (see **Gravity models of migration: Available data sets**), it is important to fully understand the type and the impact of input data used during analysis when applying gravity models. For instance, an international migrant can be defined according to different definitions depending on how we consider his/her origin. Labor market analysis usually defines immigrants based on their country of birth, as acquiring the nationality of a host country is part of the assimilation process under examination. However, most studies focusing on immigrant mobility are more interested in a migrant's last place of residence rather than their country of birth or citizenship. With that said, the decision about which origin country to consider is usually made based on data availability. Here it is important to realize that some of the determinants of migration that will be included in the gravity model will vary depending on the accepted definition of an international migrant. For instance, visa restrictions are based on citizenship rather than residence, while linguistic proximity is much more closely related to the country of birth than to the country of residence.

Measurement issues (II): Bilateral gross flows, net flows, stocks, or variation in stocks?

In the presence of migration costs, the decision to move to a particular country and the decision to stay in that country are not the same. For this reason, analysis of the determinants of migration should be based on origin-destination data (also called dyadic data) and, in particular, on bilateral gross flows (the absolute value of individuals moving from one country to the other in a particular direction). However, limitations in data availability have caused researchers to follow other alternatives (e.g. stocks, variation in stocks or net flows, or the absolute difference between emigrants and immigrants between the two analyzed countries). This approach poses challenges when using gravity models, as it is quite clear that variations in stocks are subjected to measurement errors when used as a proxy for gross flows. In fact, variations in stocks are influenced by return migration or migration to third countries, and, as a result, negative values could be obtained. The impact of negative data values is addressed in the following section.

Other researchers using migration stocks have interpreted their results as a representation of long-term equilibrium [1]. They also argue that because data on immigration stocks are usually based on national censuses, they are probably of higher quality than those sources reporting annual immigrant flows. The main reason is that censuses deal with unambiguous net permanent moves (the total number of immigrants less the number of emigrants in a particular period) and reduce the undercounting of undocumented immigrants. However, as censuses are usually carried out every ten years, they can only provide interesting insights in the medium and long term that are not compatible with the RUM model.

Logs and zeros: Implications for estimation procedures

One challenge that arises when researchers derive gravity models from RUM models by using natural logarithms is how to deal with the potential presence of zero or

negative values (in case net flows are used) for bilateral migrant flows. Regarding negative values, researchers usually exclude them from the sample or correct them as in [3]. In the case of zeros, the most common strategies are to omit these observations or to arbitrarily add a small positive number (usually 0.5 or 1) in order to ensure that the logarithm is well defined. However, by deleting zero flows, relevant information on pairs of countries where there are no migratory movements is not taken into account. Adding a positive number is also problematic, as small variations in the selected number will produce big variations in the results [4]. For these reasons, the literature is considering two alternative procedures: first, to use count data models such as Poisson, negative binomial, and zero-inflated models (see [5] for detailed descriptions of these approaches), and second, to apply Heckman's selection model in order to correct for the probability of migration in the gravity equation.

However, there are still some technical issues that should be taken into account when applying these procedures. For instance, Poisson pseudo maximum likelihood eliminates the need to use natural logarithms—thereby reducing the problems associated with zero and negative data points—but tends to over-weigh high-value flows and, moreover, the estimation can face problems of convergence toward the optimal values of the parameters. Regarding Heckman's procedure, the main difficulty is to find an appropriate instrument, a variable explaining the absence of flows, but that is not related to the size of flows. For instance, the existence or not of diplomatic representation among the considered countries has been used, as this variable might affect the probability of initial migration but not necessarily the magnitude of the flows [6]. For example, in the absence of any diplomatic representation of country A in country B, the cost of obtaining a visa could discourage citizens of country B from trying to migrate to country. The same authors have, however, shown that their results are consistent, even when they do not consider any instrument and when using the same set of variables to predict both the possibility of having a migration flow between countries and the intensity of those flows.

Fixed effects: Omitted variable bias and multilateral resistance to migration

Gravity models are typically enlarged with additional variables related to pull and push factors. However, the omitted variable bias (the negative effects on the estimation when incorrectly leaving out one or more relevant variables) is also present in this specification. Due to improved access to longer time series of bilateral flows and the use of panel data—that is, observing the same country pairs over multiple time periods—researchers are able to include a set of country dummies (variables that take the value 0 or 1 to indicate the absence or presence of a characteristic, here called country fixed effects) to control for the average differences across countries in any observable or unobservable predictors. In this setting, gravity models are also usually enlarged with time fixed effects that account for common shocks to all countries considered in the analysis.

Another issue that is relevant when specifying a gravity model for migration analysis is to consider multilateral resistance to migration. This term is related to the influence of third countries in determining migration flows between two particular countries. For example, if two countries were moved to Mars, migration flows between them

would clearly increase, due to the current lack of alternative destinations, although their relative characteristics remain unchanged [7]. Thus, not considering the influence of potential alternative destinations could bias the results of policy analysis. For instance, in the presence of some degree of coordination in migration policies between destination countries, studies that control for multilateral resistance to migration tend to find much larger effects from these policies than studies that do not control for it. Different methods to control for multilateral resistance to migration have been proposed. In case the data set has the appropriate longitudinal dimension (high number of country-pairs and time periods), the solution involves applying the Common Correlated Effects estimator that allows for the introduction of cross-sectional averages of the dependent and independent variables [8]. In case this option is not feasible due to data limitations, one possible solution is to include origin-year dummies [2] or destination-year dummies [3].

It is worth mentioning that the inclusion of dyadic fixed effects also helps alleviate other potential negative effects. For instance, gravity models implicitly assume that costs increase linearly according to (log) distance, though this is not always true, as it may be cheaper to travel further along a well-traveled route than to a less popular destination nearby. Incorporating dyadic fixed effects into the gravity model captures these factors, so long as the relative cost ranks remain similar over time (e.g. no new bridge is built between an island and the mainland that would alter the related travel costs).

Structural and policy analysis

The use of gravity models within the context of migration has shed light on how different exogenous (external) factors affect migration flows. Some of these factors are related to characteristics of the origin or destination country, such as the existence of better labor market prospects or some general immigration policies, while other factors are directly related to the particular pair of countries considered such as the existence of bilateral immigration agreements.

Studies related to the analysis of environmental factors provide an example of the first group—that being studies that deal with single country specific factors. In particular, the objective in one study was to investigate to what extent international bilateral migration flows between 1960 and 2000 could be explained by natural disasters and climatic variations [3]. The authors' analysis is oriented toward the medium- and long-term effects of climate change; they found no support for any direct relationship between climatic factors and international migration, although natural disasters did have a direct effect on internal migrations as urban environments become more attractive.

An example of the second group of studies, where gravity models are used to assess the effects of dyadic variables on migration flows, include the analysis of linguistic proximity. In one study the authors constructed a measure of linguistic proximity between origin and destination countries; they found that language affects migration costs, even after considering the effects of cultural homogeneity or physical proximity [9]. In fact, the impact of linguistic proximity on bilateral migration flows is much stronger than the impact of country differences in terms of unemployment rates,

although the effect is lower than that of ethnic networks or other traditional pull and push factors.

In this context, gravity models have also been applied to consider the impact of policies affecting migration flows between origin and destination countries. The aim is to quantify the effect of a specific policy on flows, controlling for the remaining pull and push factors. For instance, in a longitudinal data framework, the impact of bilateral policies (such as, the elimination of visa restrictions between two countries) can be considered by including a variable that allows for the identification of the policy's impact by exploiting variations over time (before and after the policy) and across countries (those affected and not affected by the policy). However, depending on data availability and the manner in which the policy is defined, this is not an easy task. We have already established that it is important to control for multilateral resistance to migration, as most policies not only have a direct effect on the flows between the two countries in question, but will also alter the relative attractiveness of alternative destinations. The method of controlling for multilateral resistance to migration (such as the inclusion of origin-year dummies or destination-year dummies) could create an identification problem between the fixed effects and the policy variable. For instance, if a particular migration policy is adopted by a country during the entire time period under consideration, then there will be a multicollinearity problem (a statistical association between different explanatory variables) in our model due to the inclusion of the policy variable and the origin-year fixed effects, which will be difficult to disentangle (e.g. if migration increases between two countries after they have signed a visa agreement, but at the same time, the economic situation worsens in the origin country compared to the destination one). A possible solution is to apply bound analysis (i.e. set minimum and maximum values for the variable in question) when analyzing the impact of visa policies in explaining bilateral migration flows [10]. Previous literature has not found any significant effects of visa policies on migration flows. However, after controlling for multilateral resistance to migration and calculating average bounds, the introduction of a visa requirement lowers incoming flows by between 40% and 47%. The policy also had some indirect effects that were not previously considered. In particular, the introduction of a visa requirement by one destination increases flows toward other countries by between 2.8% and 16.9% [10].

An additional caveat when considering policy impacts in the context of gravity models is related to the potential presence of endogeneity (i.e. countries that are more exposed to immigration could decide to adopt more restrictive policies, which creates a circular chain effect between migration and policies). In order to solve this problem, one possibility would be to apply instrumental variables estimators, although it is quite difficult to find appropriate instruments. This would require finding variables related to the policy we want to analyze, but at the same time, they should not be correlated to the rest of the regressors in our gravity model. For this reason, the use of internal instruments such as past bilateral flows does not always solve the problem, and the identification of external instruments is always difficult unless there are some historical events that can help us to identify the instrumental variable.

For example, one recent study analyzes the drivers of international students' mobility using a gravity model [11]. Reverse causality can also be a concern when interpreting the results of two of the authors' specification's regressors. First, they consider the

role of networks, proxied by the stock of educated migrants in destination countries at the beginning of the considered period. The idea is that students engaged in higher education benefit from the support of skilled migrants in the destination country. However, if the destination countries favor migration from some particular origin countries, it stands to reason that they will also favor the arrival of students from those countries. As a result, the observed positive impact of networks on migration flows will actually be the result of students simply following the general pattern of economic migrants or those affected by family reunification programs that come from the same origin countries as the students. In order to disentangle the effects of networks on flows, the researchers use the following instrument: the existence of guest worker programs after the Second World War that attracted economic migrants to work in some specific industries, like coal mines or steel factories. These guest worker agreements led to important diasporas in the destination countries, and are good independent predictors of migrant networks. The results when using this instrument still support the positive impact of networks on international student flows. The second variable upon which they apply an instrument is enrolment fees. According to their initial results, there is a positive correlation between higher fees in destination countries and higher flows of international students. Although this unexpected result could be explained as a signal for the presence of higher quality education in the destination country, it could also be related to reverse causality. Those universities that are more attractive for international students can afford to charge them higher fees. As this policy can easily be implemented by private universities, the researchers use the following instrument: the private sector's share of total expenditures in the higher education systems in destination countries, a variable that is related to the capacity of universities to charge higher fees, but not necessarily explaining international student arrivals. When using this instrument they obtain no significant effect of fees on flows, a result that could also be related to the existence of grants for international students.

It is worth mentioning that the use of instrumental variables can also be justified as a way to correct the potential bias derived from omitted variables. However, if multilateral resistance to migration has been considered by the inclusion of fixed effects, the problem is usually alleviated.

LIMITATIONS AND GAPS

The primary limitation to gravity models within the context of migration analysis has been the limited availability of bilateral migration data; however, the situation is improving quickly. Nowadays, the main concerns are related to issues such as multilateral resistance to migration or the frequent presence of zero observations. Data limitations are also more evident when the focus is not on international migration, but on internal migration, a topic that is receiving increasing attention, and where microdata from censuses are becoming the most relevant source for this kind of analysis.

Some authors have also used gravity models to estimate potential future migration flows between different pairs of countries. For instance, a 2010 study specified and estimated a gravity model that could be included as part of a demographic projection model [12]. Taking this objective into account, the authors selected explanatory variables that could easily be projected in terms of demographic scenarios. Although

this analysis can provide some insights to help predict migration flows, the use of gravity models from this perspective will provide only limited utility in assessing the impact of different policy scenarios. In particular, although gravity models could be used to perform counterfactual evaluations of the evolution of migration flows after policy changes, counterfactuals need to be performed carefully, and must properly take into account the impact of multilateral resistance to migration. As previously mentioned, one relevant assumption of the RUM model is that the attractiveness of a destination is not supposed to be affected by migration, which may not always be the case in reality. Researchers using gravity models to calculate migration potentials should be aware of this limitation.

SUMMARY AND POLICY ADVICE

The use of gravity models as a tool to analyze international migration flows has substantially increased during the last decade. The improved availability of bilateral migration data has allowed researchers to analyze the role of pull and push factors that had not previously been considered in the literature. Taking advantage of these new data sets, researchers have obtained new evidence on the role of networks, or on the relative contribution of linguistic proximity between countries, to help explain international migration flows. As a part of structural analysis, gravity models have also been used to shed new light on the impact of migration policies such as bilateral visa restrictions.

Moreover, and in a parallel way to the recent evolution of studies focusing on bilateral trade, new theoretical and methodological advances have allowed researchers to overcome some of the methodological challenges posed by the use of dyadic (i.e. bilateral origin to destination) data. For instance, by referencing the underlying theoretical framework in terms of RUM models, gravity models now allow researchers to clarify key points for applied research such as the need to control for multilateral resistance to migration. However, there are still some issues that require further attention in order to improve the applicability of gravity models such as the excess of zeros in data sets or the presence of endogeneity problems.

Acknowledgments

The author thanks two anonymous referees and the IZA World of Labor editors for many helpful suggestions on earlier drafts. Previous work of the author (together with Jordi Suriñach) contains a larger number of background references for the material presented here and has been used intensively in all major parts of this article (Ramos, R., and J. Suriñach. *A Gravity Model of Migration between ENC and EU*. IZA Discussion Paper No. 7700, 2013).

Competing interests

The IZA World of Labor project is committed to the *IZA Guiding Principles of Research Integrity*. The author declares to have observed these principles.

REFERENCES

Further reading

Beine, M., S. Bertoli, and J. Fernández-Huertas Moraga. "A practitioner's guide to gravity models of international migration." *World Economy* (Forthcoming).

Head, K., and T. Mayer. "Gravity equations, workhorse, toolkit, and cookbook." In: Helpman, E., K. Rogoff, and G. Gopinath (eds). *Handbook of International Economics*, *Volume 4*. Amsterdam: Elsevier, 2014; pp. 131–195.

Ramos, R., and J. Suriñach. *A Gravity Model of Migration between ENC and EU*. IZA Discussion Paper No. 7700, 2013.

Key references

[1] Grogger, J., and G. H. Hanson. "Income maximization and the selection and sorting of international migrants." *Journal of Development Economics* 95:1 (2011): 42–57.

[2] Ortega, F., and G. Peri. "The effect of income and immigration policies on international migration." *Migration Studies* 1:1 (2013): 47–74.

[3] Beine, M., and C. Parsons. "Climatic factors as determinants of international migration." *Scandinavian Journal of Economics* 117:2 (2015): 723–767.

[4] Burger, M., F. van Oort, and G.-J. Linders. "On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation." *Spatial Economic Analysis* 4:2 (2009): 167–190.

[5] Santos Silva, J. M. C., and S. Tenreyro. "The log of gravity." *Review of Economics and Statistics* 88:4 (2006): 641–658.

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

[7] Bertoli, S., and J. Fernández-Huertas Moraga. "Multilateral resistance to migration." *Journal of Development Economics* 102 (2013): 79–100.

[8] Pesaran, M. H. "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica* 74:4 (2006): 967–1012.

[9] Adserà, A., and M. Pytliková. "The role of language in shaping international migration." *The Economic Journal* 125:586 (2015): F49–F81.

[10] Bertoli, S., and J. Fernández-Huertas Moraga. "The size of the cliff at the border." *Regional Science and Urban Economics* 51 (2015): 1–6.

[11] Beine, M., R. Noël, and L. Ragot. "Determinants of the international mobility of students." *Economics of Education Review* 41 (2014): 40–54.

[12] Kim, K., and J. E. Cohen. "Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity." *International Migration Review* 44:4 (2010): 899–932.

[13] Ramos, R., and J. Suriñach. *A Gravity Model of Migration between ENC and EU*. IZA Discussion Paper No. 7700, 2013.

Online extras

The **full reference list** for this article is available from:
http://wol.iza.org/articles/gravity-models-tool-for-migration-analysis

View the **evidence map** for this article:
http://wol.iza.org/articles/gravity-models-tool-for-migration-analysis/map