International Migration Systems: Evidence from Harmonized Flow Data

Jack DeWaard,[†] Keuntae Kim,[†] and James Raymer[‡]

Abstract

We use harmonized data on bilateral migration flows between countries in the European Union (EU) from 2003 to 2007 to test whether international migration systems are structured according to "different hierarchies" per the expectations of migration systems theory (Zlotnik 1992:39). Migration systems theory requires evidence from data on migration flows for the determination of migration systems; yet available data of this sort lack a consistent metric given differences in national systems of data collection and with the timing criteria used to validate migrations/migrants. We use harmonized estimates of migration flows from the MIgration MOdeling for Statistical Analyses (MIMOSA) project, which combine the emigration and immigration reports of origin and destination countries into a single set of flows that are consistent and complete, to develop explanatory models estimating the size of migration flows in relation to socio-cultural, geopolitical, economic, and demographic covariates. Results from modified gravity models and fixed effects vector decomposition support two hierarchies – an *explanatory* hierarchy from the salience of factors associated with origin-destination pairs, and an *exposure* hierarchy which distinguishes the size of migration flows on the basis of the relative tenure of origin and destination countries in the EU. Our findings lend support to migration systems theory and constitute the first such evidence using harmonized data on migration flows.

[†] Department of Sociology and Center for Demography and Ecology, University of Wisconsin Madison, 8108 William H. Sewell Hall, Madison, WI 53706. Email: jdewaard@ssc.wisc.edu; ktkim@ssc.wisc.edu.
Jack DeWaard acknowledges support from NICHD Training Grant T32 HD07014 and center grant R24 HD047873 to the Center for Demography and Ecology at the University of Wisconsin Madison.
[‡] Division of Social Statistics, School of Social Sciences, University of Southampton, Southampton SO17 1BJ. Email: J.Raymer@soton.ac.uk.
*The authors acknowledge use of harmonized estimates of bilateral migration flows provided by de Beer et al. (2010) as part of the MIgration MOdeling for Statistical Analysis (MIMOSA) project.

Introduction

Migration systems theory (MST) situates international migration as a product of interacting nation-states and congruent socio-cultural, geopolitical, and economic factors and policies (Zlotnik 1992). First articulated by Mabogunje (1970) and extended by Kritz and Zlotnik (1992), MST rests on tenuous empirical footing. If "international migration were perfectly measureable, migration systems might be identified by examining the matrices of in-flows, out-flows, and net-flows between all countries as they evolved through time" (Zlotnik 1992:20). Available data of this sort, however, lack a consistent metric given differences in national systems of data collection and with the timing criteria used to validate migrations/migrants.

The "comparability of international migration statistics" is a well known and, to a lesser extent, well documented problem (Lemaitre 2005:1; Poulain et al. 2006). Contrast Sweden, for example, which collects data on *migrations* from a single population register of nationals and foreigners with the United Kingdom, which collects data on *migrants* from the International Passenger Survey. Sweden and the UK employ a *one-year* timing criterion to validate the emigrations and immigrations of nationals and foreigners. Other countries, such as Romania, maintain separate population registers for nationals and foreigners and track only their respective emigrations and immigrations. Romania also employs a *permanence* criterion, which invalidates migrations that Sweden and the UK would consider legitimate.

These discrepancies render the determination of international migration systems tenuous given the "lack of comparable data on migration flows" (Zlotnik 1992:32). That MST lacks evidence from consistent data on bilateral migration flows serves as the motivation for this project. We use harmonized estimates of migration flows between countries in the European Union (EU) from 2003 to 2007 obtained from the MIgration MOdeling for Statistical Analyses (MIMOSA) project (de Beer et al. 2010; de Beer et al. 2009; Raymer & Abel 2008; http://mimosa.gedap.be). These estimates combine the emigration and immigration reports of origin and destination countries into a single set of flows that are consistent and complete. The MIMOSA project used optimization techniques to harmonize data where both immigration and emigration reports were available. These flows were benchmarked to Sweden's immigration reports, which align with the

United Nations' (1998) recommendation for a one-year timing criterion for long-term migration. Regression models were then applied to fill in missing values. Raymer and Abel (2008) set out the general framework for these estimates, albeit with a different harmonization method than the one ultimately used by MIMOSA.

The MIMOSA estimates serve as the outcome in explanatory models used to estimate the size of migration flows in relation to a set of socio-cultural, geopolitical, economic, and demographic covariates to test whether the EU migration system is structured according to "different hierarchies" per the expectations of MST (Zlotnik 1992:39). Specifically, we posit and test for an *explanatory* hierarchy evidenced by the salience of factors associated with origin-destination pairs relative to those associated with origin and destination countries, respectively. We likewise posit and test for an *exposure* hierarchy which distinguishes the size of migration flows on the basis of the relative tenure of origin and destination countries in the EU.

In addition to being the first to use harmonized and complete data on migration flows to leverage a test of MST, we incorporate fixed effects vector decomposition (FEVD) into the explanatory framework of the modified gravity model (Plümper & Troeger 2007; Greenwood 1997). FEVD estimates the effects of key time-invariant features of origin-destination pairs at the core of MST, controlling for residual unit effects.

Theoretical and Empirical Background

Migration Systems Theory

MST is a theoretically encompassing perspective that synthesizes and supplements the accounts of international migration provided by neoclassical economics, the new economics of migration, world systems theory, bifurcated labor market theory, social capital theory, and cumulative causation. It is, according to Massey et al. (1998:60), an encompassing theory of international labor markets. At a more basic level, however, MST is a theory about interacting nation-states and embedded socio-cultural, geopolitical, and economic linkages which establish international migration at levels where migration systems obtain (Zlotnik 1992; Boyd 1989; Fawcett 1989).

These interdependencies give rise to sustained and sizeable bilateral migration flows such that the determination of a migration system can presumably be made on the basis of these largely exogenous characteristics.

Zlotnik (1992) discusses the *existence* and *properties* of migration systems. The existence of a migration system is determined by specifying a "threshold beyond which migration is considered to create a 'strong' link between nation states…[A]ny submatrix whose entries remained above the threshold during five or ten years would indicate the potential existence of a system" (Zlotnik 1992:20). Despite this opaque definition, it nonetheless implicates the aforementioned dynamic in shaping the level (i.e., threshold) of migration flows.

A stylized view of the properties of migration systems is provided by Fawcett (1989) who posits three sets of linkages – relational, regulatory, and tangible – between countries in a migration system. Zlotnik (1992:20) has similarly suggested shared geographic region, "comparable levels of development,…cultural affinity," and congruent migration policies as essential features of migration systems. Judged on these characteristics, both North America and Western Europe have been deemed international migration systems (Massey et al. 1998; Zlotnik 1992).

One must also "establish the different hierarchies that may be operating within a system" (Zlotnik 1992:39). At least two can be envisioned. In predicting the size of migration flows, an *explanatory* hierarchy is evidenced by the salience of factors associated with origin-destination pairs relative to those associated with origin and destination countries, respectively. The shared "community" factors of origin-destination pairs not only make unique explanatory contributions relative to those associated with origin and destination countries, as evidenced in the literature on gravity-based approaches to bilateral migration flows (van Tubergen et al. 2004:705), but these factors are viewed as explanatorily salient.

Differences between origin and destination countries in a migration system also suggest that an *exposure* hierarchy might be operative, which distinguishes the size of bilateral migration flows on the basis of the relative tenure of origin and destination countries in a migration system. Migration systems expand and contract on the basis of the aforementioned 5-10 year threshold

(Zlotnik 1992:20). Differential exposure to the migration system may introduce heterogeneity with respect to the size of bilateral migration flows.

Evidence for Migration Systems

Empirical scholarship on migration systems comes in two variants. *Indirect* efforts are not motivated by MST, yet rely extensively on available emigration and immigration reports of origin and destination countries. *Direct* efforts, in contrast, are motivated by MST, but are more cognizant of the inherent problems with available data on migration flows (Massey et al. 1998; Zlotnik 1992). While each has its advantages, neither is consistent with the call to examine "the matrices of in-flows, out-flows, and net-flows" using data that are "perfectly measureable" or, at a minimum, internally consistent (Zlotnik 1992:20).

Indirect efforts assume the form of gravity-based approaches reminiscent of Ravenstein's (1885) laws of migration, where migration is "proportional to the pressure gradient between [countries] and inversely proportional to the resistance" (Thornthwaite 1934:1). Migration flows between countries for which "populations are P_1 and P_2 and which are separated by…distance, D, will be proportionate to the ratio, $P_1 * P_2 / D$" (Zipf 1946:677). Taking the natural logarithm and adding a random error term, we arrive at the *gravity model*:

$$\ln(m_{ijt}) = \alpha + \beta_1 \ln(p_{it}) + \beta_2 \ln(p_{jt}) + \beta_3 \ln(d_{ij}) + e_{ijt} \qquad (1)$$

where m_{ijt} is the migration flow from country i to j at time t, p_{it} and p_{jt} are the populations of countries i and j at time t, and d_{ij} is the distance separating countries i and j.

The gravity model is convenient for explanatory purposes and used extensively in indirect efforts (Greenwood 1997). Cohen et al. (2008) analyzed the emigration and immigration reports of 11 countries between 1960 and 2004. Coefficients for the log of population at origin and destination were positive with elasticities less than 1.0, suggesting that migration flows are not especially sensitive to population size at origin and destination (Karemera et al. 2000). The coefficient for logged distance was negative, which is expected as distance is a proxy for the costs of migration.

The gravity model is extended by dropping and adding terms to (1), which Cohen et al. (2008) also did by considering population density and country fixed effects. The resulting *modified gravity model* can thus incorporate "behavioral content in the context of the gravity-model approach" (Greenwood 1997:664). Additional covariates are generally four types in practice – demographic, economic, geopolitical, and socio-cultural. We summarize each in turn.

Kim and Cohen (forthcoming) estimated separate explanatory models for the emigration and immigration reports of 13 origin and 17 destination countries between 1950 and 2007 using data from the United Nations. Noting that demographic and economic covariates are potentially endogenous, they examined population size, the relative size of the urban population (i.e., percent urban), and the potential support ratio (PSR) among others. Population size at origin and destination was positively associated with emigration and immigration flows. Supporting Neumayer's (2005) contention that those in urban areas are better informed about migration, the percent urban was positively associated with emigration flows. While labor market competition may explain why the PSR at origin increased the size of emigration flows, the corresponding negative relationship between the PSR at origin and immigration flows is opposite that reported by Leblang et al. (2009) and Mayda (2005).

Economic covariates used in indirect efforts include measures of wage rates and GDP per capita as indicators of economic advantage (Mayda 2005); the GDP per capita ratio as an indicator of relative economic advantage (Leblang et al. 2009; Greenwood & McDowell 1991); performance ratings of financial institutions and the rate of unemployment as indicators of economic risk and deprivation (Karemera et al. 2000); and rates of labor force participation and degree of industrial similarity between origin-destination pairs as indicators of economic development and labor market segmentation (Massey & Espinosa 1997; Greenwood & McDowell 1991).

Higher wages at origin reduced rates of immigration to the United States and Canada between 1962 and 1984 (Greenwood & McDowell 1991). GDP per capita at origin, however, had little effect on rates of immigration for 14 OECD countries between 1980 and 1995 (Mayda 2005). The GDP per capita ratio reduced rates of immigration to 26 OECD countries by an average of

3.4 percent between 1985 and 2004 when favoring origin countries (Leblang et al. 2009). And modest, positive associations were reported by Karemera et al. (2000) and Greenwood and McDowell (1991) between rates of immigration to the United States and Canada between the 1960s and 1980s and the performance of financial institutions, the total unemployment rate, and rates of labor force participation at origin.

Provisions of the welfare-state contain both economic and geopolitical elements. Svaton and Warin (2007) estimated the effects of social expenditures on immigration flows reported by EU-15 countries between 1994 and 2004. They found that higher social expenditures at destination increased total immigration flows and immigration flows from Central and Eastern Europe and the developing world. Reported elasticities for the latter were greater than 1.0, suggesting that immigration flows are especially sensitive to social expenditures at destination, a finding echoed by Leblang et al. (2009). The impact of immigration policy is difficult to measure; the available evidence suggests that hemispheric quotas unfavorable to immigration from the Asia-Pacific triangle prior to 1965 and to immigration from Europe between 1969 and 1976 had negative effects on the rate of immigration to the United States between 1962 and 1984 (Greenwood & McDowell 1991). The provision of political rights and civil liberties at destination has not received substantial empirical support (Leblang et al. 2009; Pedersen et al. 2008).

Socio-cultural covariates include intersecting national and colonial histories and common official language(s) of origin-destination pairs. As measures of social distance, each was reported to be positively associated with immigration flows to 27 OECD countries between 1989 and 2000 (Pedersen et al. 2008). Effect sizes generally less than .40 are robust to the estimation of separate models for emigration and immigration flows (Kim and Cohen forthcoming).

In contrast to indirect efforts, direct efforts have relied on data on migrant stocks to provide descriptive accounts of migration systems. In the 1980s, 85 percent of all immigrants to North America came to the United States (Massey et al. 1998). About 51 percent of all immigrants in the United States were from Central and Latin America by the 1990s (Martin & Midgley 2003). Recent data on migrant stocks reflect the legacy of undocumented immigration to the United States. The shift toward a policy of family reunification in the 1960s followed by the legalization of 2.7 million temporary agricultural workers with the Immigration Control and Reform and Control Act of 1986 further entrenched these linkages (Borjas 1999; Massey & Espinosa 1997).

In Western Europe, the migration of guest-workers from the European periphery after World War II was replaced by migration from Central and Eastern Europe by the 1990s (Castles & Miller 2003; Massey et al. 1998). EU expansions in 2004 and 2007 strengthened these linkages by reducing barriers to migration. Countries such as Germany have become important intermediary and final destinations for immigrants from Eastern Europe (Zlotnik 1992). As in North America, the migration system of Western Europe centers on current and former labor importing countries, suggesting the dominance of economic motivations for migration reflecting global investment patterns, the erosion of skilled manual occupations, and expansion of the service sector (Castles & Miller 2003; Piore 1979; Wallerstein 1974).

Comparability of Migration Flow Data

The distinction between *indirect* and *direct* efforts rests on being motivated by MST and the data employed. With respect to the latter, data on migrant stocks are insufficient for the determination of migration systems because they confound mortality and naturalization with migration, as well as obscure the distinction between past and recent migrants (Zlotnik 1992:32). While residual approaches can be used to estimate the net size of migration flows from the population balancing equation applied to two consecutive U.S. censuses, this approach does not address naturalization, nor is it clear how to distribute the resulting decadal flows across years.

Data on bilateral migration flows reported by origin and destination countries are therefore a more promising avenue for the determination of international migration systems. These data are not, however, fit for use in their available forms. A common way to examine the discrepancies between the emigration and immigration reports of origin and destination countries is to construct a double-count matrix. Figure 1 is adapted from Kupiszewska and Nowok (2008) and presents such a matrix using data on migration flows in 2003 classified by next/previous country of residence for EU-15 countries obtained from Eurostat's New Cronos database.

FIGURE 1 ABOUT HERE

Figure 1 presents 420 possible emigration and immigration reports of origin and destination countries on what are 210 actual migration flows (210 = 15*15-15). Row and column identifiers denote origin and destination countries. Each element of the matrix has three values; the top two are the emigration and immigration reports of origin and destination countries, respectively, and the third is the ratio of reported emigration to immigration when both are non-missing.

Two observations can be made. First, the data are substantially incomplete. Second, with the exception of Nordic countries, the emigration and immigration reports of origins and destinations are seldom in agreement (i.e., $E/I = 1$). Moreover, these discrepancies are not unique to the particular cross-section. Figure 2 displays the emigration and immigration reports for selected origin and destination countries in the EU between 2003 and 2007 using data from Eurostat. Near perfect agreement between Sweden and Denmark can be contrasted with differences in the reported levels (Germany to Finland) and stability (UK to the Netherlands) of migration flows.

FIGURE 2 ABOUT HERE

These discrepancies reflect differences in national systems of data collection and with the timing criteria used to validate migrations/migrants. A population registration system for nationals and foreigners can be compared against any number of alternative arrangements – e.g., separate population registration systems for nationals and foreigners in the case of Slovenia, use of residence permits for foreigners in the case of Hungary, household-based and border surveys by Ireland and the United Kingdom, and use of other administrative data. The timing criteria used to validate migrations/migrants also vary across countries. The Untied Nations (1998) advocates a 1-year timing criterion; in practice, however, timing criteria cover the spectrum from none to permanence, with 3-, 6- and 12-month variants between. A taxonomy of these conventions for countries in Europe can be found in Poulain et al. (2006).

The lack of consistent and complete data on bilateral migration flows raises questions about their use in empirical work. In the current project, we use *harmonized* estimates of migration flows between countries in the EU between 2003 and 2007 obtained from the MIMOSA project, the details of which are discussed in the following section. The MIMOSA estimates serve as the outcome in explanatory models used to estimate the size of bilateral migration flows from socio-cultural, geopolitical, economic, and demographic covariates to test whether the EU migration system is structured according to "different hierarchies" per the expectations of MST (Zlotnik 1992:39). Specifically, we assess evidence for two forms of hierarchy – an *explanatory* hierarchy from the salience of factors associated with origin-destination pairs and an *exposure* hierarchy which distinguishes the size of bilateral migration flows on the basis of the relative tenure of origin and destination countries in the EU.

Data and Methods

Harmonized Migration Flows

Harmonized data on bilateral migration flows were obtained from the MIMOSA project (de Beer et al. 2010; de Beer et al. 2009; Raymer & Abel 2008). The MIMOSA method consists of two steps. The first step harmonizes data for countries that provide both immigration and emigration data by country of origin and destination, respectively. In total, 19 countries provided such data from 2002 to 2007. MIMOSA used optimization techniques to calculate adjustment factors for emigration and immigration benchmarked against Sweden's migration data. Sweden provides reliable migration data and reports are consistent with the United Nations' (1998) recommended definition for long-term migrants.

MIMOSA's second step was to estimate migration flows for countries that do not provide data. Coefficients from pooled regression models using the harmonized data (first step) were used to estimate missing immigration and emigration totals within and outside the EU, as well as origin-destination associations within the matrix. Regression equations for emigration and immigration totals included variables such as population size and GDP per capita. For the missing associations, variables such as distance, contiguity, migrant stocks, and language family were used. Refer to Raymer and Abel (1998) for more detail on the estimation of missing data.

In the current project, we employ the MIMOSA estimates of migration flows between countries in the EU from 2003 to 2007. We limit our sample to intra-EU migration flows for theoretical reasons. Zlotnik (1992) distinguished between the existence and properties of migration systems and the requirements for demonstrating each. Since determining both would require more space than we have, we focus on the latter and assume the existence of an EU migration system via our sample and on the basis of existing work to date (Massey et al. 1998; Zlotnik 1992). We analyze 2,712 migration flows between origin-destination pairs in the EU from 2003 to 2007. The data constitute an unbalanced panel given EU expansions in 2004 and 2007 which increased the country set from 15 to 25 and from 25 to 27 (2,712 = (15*15-15)+3*(25*25-25)+(27*27-27)).

Covariates

A list of covariates, descriptions, and data sources is provided in Table 1. We use a continuous term for time, centered at 2003. Data on the geographic distance (in kilometers) between the largest cities of origin-destination pairs is from Centre D'Etudes Prospectives et D'Informations Internationales (CEPII). A measure of linguistic distance, we use a dummy indicator for whether two countries composing an origin-destination pair share one or more official languages (Kim & Cohen forthcoming; Pedersen et al. 2008; Mayda 2005; Karemera et al. 2000).

TABLE 1 ABOUT HERE

Demographic covariates are taken from the United Nations Population Division's *demobase*, a clearinghouse for population estimates and projections. These estimates are based on medium variant estimates and projections by the UN. Data employed include the relative size of the urban population (i.e., percent urban) and the potential support ratio (PSR) at both origin and destination (Kim & Cohen forthcoming; Leblang et al. 2009; Pedersen et al. 2008; Mayda 2005; Greenwood & McDowell 1991). For each year, the PSR is calculated as the ratio of persons ages 15-64 to persons 65 and older, multiplied by 100. The demobase provides only quinquennial estimates of the PSR, which we linearly interpolate to arrive at annual estimates.

Economic covariates include the ratio of GDP per capita (based on purchasing power parity) in constant U.S. dollars at destination to that at origin taken from the International Monetary Fund's *World Economic Outlook 2008* (Leblang et al. 2009; Greenwood & McDowell 1991). We also consider rates of total unemployment at both origin and destination taken from the International Labor Organization's (ILO) Key Indicators of the Labor Market (KILM) (Leblang et al. 2009; Svaton & Warin 2007; Karemera et al. 2000). Finally, we consider the percent change in rates of labor force participation at origin and destination taken from the ILO as a measure of economic growth and development (Svaton & Warin 2007; Neumayer 2005; Massey & Espinosa 1997; Greenwood & McDowell 1991)

From Leblang et al. (2009) and Svaton and Warin (2007), we use total government expenditures on social protection per household head at origin and destination obtained from Eurostat as one of two geopolitical measures. Since countries in the EU are bound by the Charter of Fundamental Rights, there is too little variation in political rights and civil liberties scores provided by Freedom House to warrant inclusion in the current project (Pedersen et al. 2008). Total social expenditures are thus intended to tap the broader institutional setting (Esping-Andersen 1999). We also look to tap more historically entrenched geopolitical interdependencies of origin-destination pairs. We construct a four category variable indicating the relative tenure of countries composing an origin-destination pair in the EU. The reference is that neither origin nor destination was a founding member of the EU with the creation of the European Economic Community (EEC) in 1957.

Finally, we consider a single socio-cultural covariate. Prior empirical work has examined the role of former and current colonial relationships between origin-destination pairs (Kim & Cohen forthcoming; Pedersen et al. 2008; Svaton & Warin 2007; Mayda 2005; Neumayer 2005). We do likewise, but also consider the role of shared national origins – i.e., whether countries composing an origin-destination pair were ever the same country. Using data from CEPII, we combine the above two measures into a single dummy indicator for whether countries composing an origin-destination pair were ever the same country *or* were ever in a colonial relationship, the latter defined as sharing a long period of substantial participation in governance.

Of these, three are especially relevant to the expectations of MST. These include: the GDP per capita ratio, the relative tenure of origins and destinations in the EU, and shared national or colonial origins. Each of these measures something about the origin-destination pair and thus the interdependencies between origin and destination countries. The GDP per capita ratio measures relative economic advantage. The relative tenure of origin and destination countries in the EU taps geopolitical interdependencies. And shared national or colonial origins captures residual socio-cultural linkages. Together, these measures help to glimpse the underlying dynamics of the EU migration system.

Explanatory Models

The current project employs the following modified gravity model:

$$\ln(m_{ijt}) = \alpha + \sum_{k=1}^{K} \beta_k \ln(x_{it-1,k}) + \sum_{k'=1}^{K'} \beta_{k'} \ln(x_{jt-1,k'}) + \sum_{k''=1}^{K''} \beta_{k''} \ln(x_{ijt-1,k''}) + \sum_{m=1}^{M} \beta_m z_{ij,m} + u_{ij} + e_{ijt} \quad (2)$$

where m_{ijt} is the migration flow from country i to j at time t, $x_{it-1,k}$ and $x_{jt-1,k'}$ are time-varying measures at origin i and destination j, respectively, at time $t-1$, $x_{ijt-1,k''}$ are time-varying measures for the origin-destination pair ij at time $t-1$, and $z_{ij,m}$ are time–invariant measures for the origin-destination pair ij.

This model has the same general structure as that in equation 1, with three exceptions. First, we allow for unit-specific intercepts, $\alpha+u_{ij}$. Second, we distinguish the between the time-varying and time-invariant characteristics of origins, destinations, and origin-destination pairs. Finally, since one can envision a non-recursive relationship between the size of migration flows and the demographic characteristics of origins and destinations, for example, we lag covariates in the first three product terms by one year (Pedersen et al. 2008; Mayda 2005).

We seek to eliminate the unit fixed effects, u_{ij}, prior to estimating the model in equation 2 given that we do not wish to model these effects explicitly and are unwilling to assume that the unit effects, as a random variable, are uncorrelated with all covariates (Frees 2004; Halaby 2004). We therefore proceed with the following fixed effects transformation (Wooldridge 2006):

$$
\begin{aligned}
\ln(m_{ijt}) - \overline{\ln(m_{ij})} &= (\alpha - \alpha) + \beta_k \sum_{k=1}^{K} [\ln(x_{it-1,k}) - \overline{\ln(x_{i,k})}] + \beta_{k'} \sum_{k'=1}^{K'} [\ln(x_{jt-1,k'}) - \overline{\ln(x_{j,k'})}] \\
&\quad + \beta_{k''} \sum_{k''=1}^{K''} [\ln(x_{ijt-1,k''}) - \overline{\ln(x_{ij,k''})}] + \beta_m \sum_{m=1}^{M} (z_{ij,m} - z_{ij,m}) + (u_{ij} - u_{ij}) + (e_{ijt} - \bar{e}_{ij}) \\
&= \beta_k \sum_{k=1}^{K} \ln(\overset{**}{x}_{it-1,k}) + \beta_{k'} \sum_{k'=1}^{K} \ln(\overset{**}{x}_{jt-1,k'}) + \sum_{k''=1}^{K''} \beta_{k''} \ln(\overset{**}{x}_{ijt-1,k''}) + \overset{**}{e}_{ijt} \qquad (3) \\
&= \ln(\overset{**}{m}_{ijt})
\end{aligned}
$$

This transformation eliminates the unit fixed effects and does so at the expense of estimating the effects of the time-invariant characteristics of origin-destination pairs, β_m. This is problematic because the β_m vector contains the effects of the aforementioned geopolitical and socio-cultural interdependencies of origin-destination pairs at the core of MST, namely – the relative tenure of countries composing an origin-destination pair in the EU and shared national or colonial origins. At present, we cannot estimate these effects.

A key innovation in this project is incorporation of fixed effects vector decomposition (FEVD) into the explanatory framework of the modified gravity model. Developed by Plümper and Troeger (2007), FEVD uses a three stage procedure to decompose the unit fixed effects and estimate the effects of both time-invariant and rarely-changing covariates, along with residual unit fixed effects, via pooled OLS. In the first stage, a fixed effects model is estimated excluding the time-invariant and rarely-changing characteristics of origin-destination pairs. In the second stage, the estimated unit fixed effects are recovered and regressed on the time-invariant and rarely-changing characteristics of origin-destination pairs. In the final stage, a pooled OLS model is estimated, which includes all time-varying and time-invariant covariates, along with the residual unit fixed effects from the previous stage.

To illustrate this process, we begin with equation 3 and generate the estimated unit effects:

$$\hat{u}_{ij} = \overline{\ln(m_{ij})} - \sum_{k=1}^{K} \beta_k^{FE} \overline{\ln(x_{i,k})} - \sum_{k'=1}^{K'} \beta_{k'}^{FE} \overline{\ln(x_{j,k'})} - \sum_{k''=1}^{K''} \beta_{k''}^{FE} \overline{\ln(x_{ij,k''})} - \bar{e}_{ij} \tag{4}$$

The estimated unit effects are not equivalent to the unobserved unit effects, u_{ij}, in the population. In addition to the unobserved unit effects, u_{ij}, the estimated unit effects include the effects of the time-invariant covariates, a constant term, and the mean effects of the time-varying covariates.

In the second stage, the estimated unit effects are decomposed into explained and unexplained components:

$$\hat{u}_{ij} = \sum_{m=1}^{M} \beta_m z_{ij,m} + h_{ij} \tag{5}$$

where $z_{ij,m}$ are time-invariant measures for the origin-destination pair ij, and h_{ij} are the residual unit effects.

In the third stage, a pooled OLS model is estimated using the covariates from equation 2 and the residual unit effects from equation 5:

$$\ln(m_{ijt}) = \alpha + \sum_{k=1}^{K} \beta_k \ln(x_{it-1,k}) + \sum_{k'=1}^{K'} \beta_{k'} \ln(x_{jt-1,k'}) + \sum_{k''=1}^{K''} \beta_{k''} \ln(x_{ijt-1,k''}) + \sum_{m=1}^{M} \beta_m z_{ij,m} + \gamma h_{ij} + e_{ijt} \tag{6}$$

The effects of time-invariant and rarely-changing characteristics of origin-destination pairs, β_m, are estimable, controlling for the residual unit effects, γh_{ij}. Simulations by Plümper and Troeger (2007) demonstrate that FEVD produces more reliable estimates than mixed effects models in estimating the effects of time-invariant covariates. For rarely-changing covariates, FEVD is best suited when (i) between-to-within variance is not large and (ii) rarely-changing covariates are not highly correlated with the unit effects. Standard errors are calculated using the fixed effects demeaned variance-covariance matrix, reducing the OLS degrees of freedom by the number of units. Serial correlation is controlled by specifying an autoregressive covariance structure (AR1) using a Cocrane-Orcutt transformation in equation 6.

Results

Harmonized Migration Flows

In Figure 3, we display the harmonized MIMOSA estimates of migration flows for the same origin-destination pairs displayed in Figure 2.

FIGURE 3 ABOUT HERE

The harmonized estimates more closely align with the flows reported by countries with better quality data. Harmonized flows from Germany to Finland both parallel and reflect the level of immigration reported by Finland; harmonized flows from the UK to the Netherlands likewise parallel and reflect the level of immigration reported by the Netherlands. While it may seem on the basis of harmonized flows from Germany to Finland and from the UK to the Netherlands that harmonized flows tend to fall somewhere between the emigration and immigration reports of origin and destination countries and thus constitute some sort of weighted average, this is not necessarily the case. Depending on the size of the emigration and immigration adjustment factors, harmonized flows are not bound by the reports of origin and destination countries.

Thus far we have resisted providing a general snapshot of migration flows within the EU and have focused our discussion on relevant theoretical and methodological issues. We break precedent for the moment and provide in Figure 4 a snapshot of the top ten emigration and immigration country-years in our data.

FIGURE 4 ABOUT HERE

The data used for Figure 4 sums harmonized emigration and immigration flows to or from all EU countries each year. Germany and Poland lead the way with respect to total emigrations. Germany and the UK likewise top total immigration flows from EU countries.

Descriptive Statistics

Descriptive statistics are provided in Table 2. We display annual and pooled means and standard deviations, as well as the proportion of variance due to the differences between origin-destination pairs. Means displayed are untransformed.

TABLE 2 ABOUT HERE

The mean migration flow fell between 2003 and 2004, increased between 2004 and 2006, and continued rising in 2007. The decline between 2003 and 2004 could reflect a real decline and/or the addition of 10 countries to the EU in 2004. There is wide variation in the size of bilateral migration flows, much of which is due to differences between origin-destination pairs. That the distribution of migration flows is non-normal motivates use of a log transformation in analyses.

Measures of geographic and linguistic distance show the effects of EU expansion. Mean geographic distance between the largest cities of countries composing an origin-destination pair increased between 2003 and 2004 and then again between 2006 and 2007. This is not surprising – of the ten countries admitted to the EU in 2004, nearly all were from Central and Eastern Europe. Linguistic distance – i.e., the proportion of origin-destination pairs sharing one or more official languages – follows a similar trend highlighting the growing diversity of the EU.

Mean percent urban declined only slightly with EU expansions, while the amount of between variation remained roughly constant. The percent urban is a clear example of what Plümper and Troeger (2007) consider a rarely-changing covariate – one where the ratio of between-to-within variation exceeds 2.8. Here, 98 percent of the variation in percent urban is due to the differences between origin-destination pairs. The PSR is also a rarely-changing variable; however, its most striking feature is the increase between 2004 and 2005 and subsequent decline between 2005 and

2007, due perhaps to the infusion of “younger” countries into the EU in 2004 counteracted by the continued aging of populations in Northern and Western Europe.

The GDP per capita ratio favors destination countries in 46 percent of cases, with considerable differences between origin-destination pairs. The mean total unemployment rate shows relatively less variation between origin-destination pairs. Trends in the mean total unemployment rate and the percentage change in labor force participation worsened in 2004 with EU expansion. By 2007, the mean total unemployment rate was approaching its 2003 level. By contrast, the mean percentage change in labor force participation continued to slow.

Following a jump between 2003 and 2004, mean government expenditures on social protection increased through 2007. With respect to the joint tenure of countries composing an origin-destination pair in the EU, trends mirror EU expansion as they should. The proportionate share of founding EU countries declined between 2003 and 2004 and again between 2006 and 2007. By 2006, nearly two-thirds of countries in the EU were not founding countries, a testament to the considerable expansion of the EU over such a short period of time.

The proportion of origin-destination pairs with shared national or colonial origins is small, .05-.07. EU expansions in 2004 increased these linkages and in 2007 decreased them.

Fixed Effects Vector Decomposition

FEVD models presented in Tables 3 and 4 were estimated using an autoregressive covariance structure (AR1). We note that this resulted in the loss of 702 observations given the application of the Cocrane-Orcutt transformation in equation 6. At the outset, we conducted a Durbin-Watson test to eliminate autocorrelation as a potential source of bias. The *d*-statistic was 0.502 ($d_{(19,2712)} = 0.502$). Comparing our result against the critical values provided in Savin and White (1977), we were unable to reject the null hypothesis of no autocorrelation.

FEVD is optimal when the between-to-within variance of time-varying covariates is not large. Plümper and Troeger (2007) suggest a threshold of 2.8. Examining these ratios for the time-varying covariates in our data, these ranged from 3 in the case of the GDP per capita ratio to 41 in the case of the percent urban. We therefore treat each time-varying covariate in this analysis as rarely-changing and include these in equation 5 en route to obtaining the residual unit effects, h_{ij}.

In Table 3, we present the results of baseline OLS and FEVD models. These results provide an initial glimpse at the directions and magnitudes of the coefficients relevant to the expectations of MST, namely – the GDP per capita ratio, the relative tenure of origins and destinations in the EU, and shared national or colonial origins. The signs and magnitudes of the coefficients are consistent across Models 1 and 2. The standard errors in Model 2 are calculated using the fixed effects demeaned variance-covariance matrix and are more precise. Model fit as indicated by the root mean square error is superior in Model 2. The discrepancy between the two models is due to the inclusion of the residual unit effects in the FEVD model.

TABLE 3 ABOUT HERE

In Table 4, we present results from three FEVD models. In addition to demographic covariates, Model 3 includes controls for time, geographic distance, and linguistic distance. Each year of time is associated with a 7 percent increase in the size of bilateral migration flows, on average. The coefficient for the natural log of distance is large and negative and is consistent with previous empirical work, which suggests that distance serves as a proxy for the costs of migration (Kim & Cohen forthcoming; Cohen et al. 2008; Greenwood 1997). Positive linguistic distance – i.e., origin-destination pairs sharing one or more official languages – likely lowers the psychic and monetary costs of migration, and is positively associated with the size of bilateral migration flows.

TABLE 4 ABOUT HERE

Demographic covariates include percent urban and PSR at origin and destination. Coefficients for percent urban at origin and destination are in the expected directions; the magnitude of the former aligns with that reported by Kim and Cohen (forthcoming) and Greenwood and McDowell (1991). The elasticity for percent urban at destination is unitary at 1.002, suggesting neither the sensitivity nor insensitivity of migration flows to percent urban at destination. The coefficients for the PSR at origin and destination are identical; a 1 percent increase in the PSR is associated with about a 3 percent reduction in the size of bilateral migration flows, on average. While the magnitudes of these coefficients are higher than those recorded in recent empirical work, the direction of the coefficients is consistent with Kim and Cohen (forthcoming).

In Model 4, we include economic covariates alongside the controls and demographic covariates included in Model 3. The inclusion of economic covariates slightly reduces the coefficients for the time trend and geographic distance; however, these remain largely consistent with those in Model 3. While the percent urban at origin and the PSR at origin and destination are unaffected by the inclusion of economic covariates, percent urban at destination appears more sensitive. In the presence of economic controls, the elasticity falls to 0.397.

The coefficient for the GDP per capita ratio is positive and statistically significant; bilateral migration flows between origin-destination pairs with a GDP per capita ratio favoring destination countries see migration flows about a 23 percent higher than those between origin-destination pairs where the ratio favors origin countries, on average. This finding is consistent with Leblang et al. (2009). Elasticities for the total unemployment rate at origin and destination are 1.221 and 1.363, respectively, suggesting that bilateral migration flows respond to economic uncertainty at home *and* abroad. The percent change in labor force participation at origin and destination are large and similarly positive, suggesting that economic development has played a considerable role in migration flows into the 21st century (Greenwood & McDowell 1991).

In Model 5, we include all remaining geopolitical and socio-cultural covariates. The introduction of these covariates decreases the coefficients for the time trend and linguistic distance; the effect of geographic distance becomes less negative. The coefficient for percent urban at destination is no longer statistically significant. Percent urban at origin, however, retains statistical significance, providing some support for Neumayer’s (2005) contention that those in urban areas are better informed about migration. Coefficients for the PSR at origin and destination settle around -1.8–2.0. These elasticities demonstrate the role of population aging in the EU, as a larger proportionate share of younger persons at origin and destination inhibits migration flows. With inclusion of the geopolitical and socio-cultural covariates in Model 5, effect sizes weaken for the GDP per capita ratio and the percent change in labor force participation at origin and destination, while they become more pronounced for the total unemployment rate at origin and destination.

In line with Svaton and Warin (2007), the coefficient for total social expenditures at origin and destination is positive and comparable to what these authors reported, 0.619. The relative tenure of origin and destination countries in the EU broadly taps the geopolitical interdependencies of origin-destination pairs. Relative to origin-destination pairs where neither origin nor destination was a founding EU member, migration flows between founding EU members are about 1 percent higher, on average. Migration flows between non-founding and founding members and between founding and non-founding members of the EU are about 81 percent higher, on average. Shared national or colonial origins is positively related to bilateral migration flows, raising the average level of migration flows by 103 percent.

Taking stock of model fit as measured by the root mean square error, model fit improves only slightly over Models 3-5. Relative to OLS models, e.g., Model 1, the good fit of these models is due to the inclusion of the residual unit effects, γh_{ij}, which capture the substantial differences between origin-destination pairs (see Table 2). Overall, Model 5 (and Models 3 and 4), is fairly stable across different specifications, producing results consistent with prior empirical work.

Discussion

The aim of this project is to use harmonized estimates of migration flows to leverage a test of MST, specifically whether migration systems are structured according to "different hierarchies" (Zlotnik 1992:39). While the direction and magnitude of each of the origin and destination covariates in Models 3-5 are consistent with prior research, three covariates are particularly relevant to the primary aim of this project. These include: the GDP per capita ratio, the relative tenure of origin and destination countries in the EU, and shared national or colonial origins. Each measures something about the origin-destination pair and thus the interdependencies between origin and destination countries. We earlier conceptualized an *explanatory* hierarchy evidenced by the salience of factors associated with origin-destination pairs and an *exposure* hierarchy which distinguishes the size of bilateral migration flows on the basis of the relative tenure of origin and destination countries in the EU.

With respect to the former, coefficients for the GDP per capita ratio, the relative tenure of origin and destination countries in the EU, and shared national or colonial origins are in the expected directions in Models 3-5. To some extent, the embedded economic, geopolitical, and socio-cultural linkages between origin-destination pairs help to establish the level of migration flows (Zlotnik 1992; Boyd 1989; Fawcett 1989). To gauge the extent of this dynamic, we begin by generating predicted migration flows from Model 5 with each of the three covariates relevant to MST set to one and all others are set at their means. We compare these values against a baseline set of predictions with each of the three covariates set to zero and all others set at their means. Taking the ratio of the two sets of predicted values, we can gauge the contribution of these three covariates to the overall level of migration. These ratios are displayed in Figure 5.

FIGURE 5 ABOUT HERE

Across EU tenure categories, the ratios range from 3.10 to 6.96. Migration flows between origin-destination pairs with economic, geopolitical, and socio-cultural interdependencies are larger than those between origin-destination pairs without these linkages by factors ranging from 3.10 to 6.96, depending on joint EU tenure. This is suggestive of an *explanatory* hierarchy evidenced by the contribution of features associated with origin-destination pairs. Establishing exactly what these features buy us with respect to the size of migration flows is important. Prior empirical work has treated these features as merely “initial conditions and legacies” controlled for by the inclusion of unit fixed effects (Andrienko & Guriev 2004). And, while attempts to model these features explicitly are not uncommon, the strategies employed make use of the methods and assumptions of mixed effects models (Kim & Cohen forthcoming; Pedersen et al. 2008; Mayda 2005), a track we sought to avoid in this analysis.

With respect to our conceptualization of an *exposure* hierarchy, comparing the ratios in Figure 5 across the categories of EU tenure, a clear difference emerges between joint founding and joint non-founding EU countries. Migration flows between origin-destination pairs with economic and socio-cultural interdependencies *and* where each country was a founding EU member are more than twice the size of migration flows between origin-destination pairs who were not founding EU members. Differential exposure of countries in the EU migration system thus appears to be a relevant explanatory factor in the size of migration flows. Perhaps somewhat counterintuitive are the ratios for origin-destination pairs not sharing joint EU tenure. One might expect that the size of migration flows between countries where the destination and not origin was a founding EU member would be larger than flows between countries where the origin and not destination was a founding EU member. Figure 5 suggest otherwise. Indeed, the opposite appears to be the case and is consistent with Martin and Taylor's (1996) "migration hump," suggesting that exposure to the migration system itself carries an institutionalizing effect, promoting migration flows of greater magnitudes when the origin (i.e., sending) country is a founding EU member.

To summarize, we find evidence for both an *explanatory* hierarchy from the salience of factors associated with origin-destination pairs, and an *exposure* hierarchy which distinguishes the size of bilateral migration flows on the basis of the relative tenure of origin and destination countries in the EU. Zlotnik (1992:39) used the phrase, "different hierarchies," to suggest an ordering to migration systems; our results are certainly consistent with this expectation.

Conclusion

We used harmonized data on bilateral migration flows between countries in the EU between 2003 and 2007 to test whether international migration systems are structured according to "different hierarchies" per the expectations of MST (Zlotnik 1992:39). We noted and described the inherent problems with the available data on migration flows. There is no common metric for migration flows given differences in national systems of data collection and with the timing criteria used to validate migrations/migrants. We therefore employed harmonized estimates of migration flows from the MIMOSA project and used these to develop explanatory models to estimate the size of bilateral migration in relation to socio-cultural, geopolitical, economic, and demographic covariates. Incorporating FEVD into the framework of the modified gravity model allowed us to examine the evidence for the aforementioned hierarchies per MST – an *explanatory* hierarchy evidenced by the salience of factors associated with origin-destination pairs and an *exposure* hierarchy which distinguishes the size of bilateral migration flows on the basis of the relative tenure of origin and destination countries in the EU. We found evidence for both forms. Our findings thus lend support to MST and constitute the first such evidence using harmonized data on migration flows.

As the current paper is a working paper, subsequent versions will include development of an original set of harmonized estimates of migration flows between countries in the EU for the period 2003-2007. van der Erf and van der Gaag (2007) and Raymer and Abel (2008) detailed an iterative method for harmonizing emigration and immigration reports of origin and destination countries based on privileging the reports of Nordic countries and applying a set of emigration and immigration adjustment factors to the reports of non-Nordic countries. This method is premised on ordering countries in the observed emigration and immigration matrices according to known data quality, which is never perfectly known *a priori*. We envision development of a procedure to take into account this uncertainty toward generating a consistent and complete set of harmonized migration flows.

Subsequent versions of this paper will also include comparison of the results from FEVD models against those from mixed effects models (e.g., Generalized Estimating Equations). The latter have been used extensively in explanatory efforts of the sort undertaken here (Kim & Cohen forthcoming; Pedersen et al. 2008).

References

Andrienko, Y. & Guriev S. 2004. "Determinants of Interregional Mobility in Russia." *Economics of Transition* 12(1):1-27.

Bauer, T. K., & Zimmermann, K. F. 1999. "Assessment of Possible Migration Pressure and its Labour Market Impact Following EU Enlargement to Central and Eastern Europe." IZA Research Report No. 3.

Borjas, G. J. 1999. *Heaven's Door: Immigration Policy and the American Economy*. Princeton, NJ: Princeton University Press.

Boyd, M. 1989. "Family and Personal Networks in International Migration: Recent Developments and New Agendas." *International Migration Review* 23(3):638-670.

Castles, S., & Miller, M. 2003. *The Age of Migration: International Population Movements in the Modern World (3rd ed.)* The Guilford Press: New York.

Cohen, J.E., M. Roig, D.C. Reuman, and C. GoSwilt. 2008 "International Migration Beyond Gravity: A Statistical Model for Use in Population Projections." *PNAS*, 105:15269-15274.

de Beer J, Raymer J, van der Erf, R., & van Wissen, L. 2010. "Overcoming the problems of inconsistent migration data: A new method applied to flows in Europe." Draft paper, Netherlands Interdisciplinary Demographic Institute, The Hague.

de Beer, J., van der Erf, R., & Raymer, J. 2009. "Estimates of OD matrix by broad group of citizenship, sex and age, 2002-2007." MIMOSA Deliverable 9.2 B, Netherlands Interdisciplinary Demographic Institute, The Hague.

Esping-Andersen, G. 1999. *Social Foundations of Postindustrial Economies*. Oxford: Oxford University Press.

Fawcett, J. T. 1989. "Networks, Linkages, and Migration Systems." *International Migration Review*, 23:671-680.

Frees, E. W. 2004. *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge: Cambridge University Press.

Greenwood, M. J. 1997. "Internal Migration in Developed Countries." In M R. Rozenzweig and O. Stark. *Handbook of Population and Family Economics*. Vol. 2. Elsevier: New York.

Greenwood, M. J., and McDowell, J. M. 1991. "Differential Economic Opportunity, Transferability of Skills, and Immigration to the United States and Canada. *The Review of Economics and Statistics*, 73: 612-623

Halaby, C. N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507-544.

Head, K. & Ries, J. 1998. "Immigration and Trade Creation: Econometric Evidence from Canada." *The Canadian Journal of Economics*, 31:47-62.

Karemera, D., Oguledo, V. I., & Davis, B. 2000. "A Gravity Model Analysis of International Migration to North America." *Applied Economics*, 32:1745-1755.

Kim, K., & Cohen, J. E. Forthcoming. "Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach Beyond Gravity." *International Migration Review*.

Kritz, M. M. & Zlotnik, H. 1992. "Global Interactions: Migration Systems, Processes, and Policies." In *International Migration Systems: A Global Approach*. Edited by M.M. Kritz, L.L. Lim, and H. Zlotnik. Clarendon Press: Oxford.

Kupiszewska, D., & Nowok, B. 2008. “Comparability of Statistics on International Migration Flows in the European Union.” In J. Raymer and F. Willekens (eds.) *International Migration in Europe. Data, Models and Estimates*. John Wiley: Chichester.
Leblang, D. A., Fitzgerald, J., & Teets, J. 2009. “Defying the Law of Gravity: The Political Economy of International Migration. Working Paper, Department of Political Science, University of Colorado Boulder.
Lemaitre, G. 2005. *The Comparability of International Migration Statistics*. Paris: OECD (Statistics Brief 9).
Mabogunje, A. L. 1970. “A Systems Approach to a Theory of Rural-Urban Migration.” *Geographic Analysis*, 2:1-18.
Martin, P., & Midgley, E. 2003. “Immigration: Shaping and Reshaping America.” *Population Bulletin* 58, no. 2 Washington, DC: Population Reference Bureau.
Martin, PL & Taylor, JE. 1996. The anatomy of a migration hump, in JE Taylor (ed), *Development Strategy, Employment, and Migration: Insights from Models*, pp 43 – 62. Paris: OECD Development Centre.
Massey, D. S., Arango, J., Hugo, G., A. Kouaouci, Pellegrino, A., & Taylor, J. E. 1998. *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Clarendon Press: Oxford.
Massey, D. S., & Espinosa, K. 1997. “What’s Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis.” *American Journal of Sociology* 102:939-999.
Mayda, A. M. 2005. “International Migration: A Panel Data Analysis of Economic and Non-Economic Determinants.” IZA Discussion Paper No. 1590.
Neumayer, E. 2005. “Bogus Refugees? The Determinants of Asylum Migration to Western Europe.” *International Studies Quarterly* 49: 389-409.
Pedersen, P. J., Pytlikova, M., & Smith, N. 2008. “Selection and Network Effects – Migration Flows into OECD Countries 1990-2000.” *European Economic Review* 52:1160-1186.
Piore, M. J. 1979. *Birds of Passage: Migrant Labor in Industrial Societies*. Cambridge: Cambridge University Press.
Plümper, T., & Troeger, V. E. 2007. “The Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Effects.” *Political Analysis* 15(2): 124-139.
Poulain M., Perrin, N., & Singleton, A. eds. 2006. *THESIM: Towards Harmonized European Statistics on International Migration.* Louvain-la-Neuve: UCL Presses Universitaires de Louvain.
Rabe-Hesketh, S., & Skrondal, A. 2008. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press: College Station, Texas.
Ravenstein, E. G. 1885. “The Laws of Migration.” *Journal of the Statistical Society of London* 48(2):167-235.
Raymer, J. & Abel, G. 2008. “The MIMOSA Model for Estimating International Migration Flows in the European Union.” Joint UNECE/Eurostat Work Session on Migration Statistics. Geneva, Switzerland. March 3-5.
Savin, N. E., & White, K. J. 1977. “The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors.” *Econometrica* 45(8): 1989-1996.
Svaton, P., & Warin, T. 2007. “European Migration: Welfare Migration or Economic Migration?” Working Paper, International Trade and Finance Association.

Thornthwaite, C. W. 1934. *Internal Migration in the United States*. University of Pennsylvania Press: Philadelphia.

United Nations. 1998. “Recommendations on statistics of international migration.” Statistical Papers Series M, No. 58, Rev. 1. Statistics Division, Department of Economic and Social Affairs, United Nations, New York.

van der Erf, R. & van der Gaag N. 2007. “An Iterative Procedure to Revise Available Data in the Double Entry Migration Matrix for 2002, 2003 and 2004.” Netherlands Interdisciplinary Demographic Institute. The Hague. http://mimosa.gedap.be/Documents/Erf.2007

van Tubergen, F., Maas, I., & Flap, H. 2004. “The Economic Incorporation of Immigrants in 18 Western Societies: Origin, Destination, and Community Effects.” *American Sociological Review* 69(5):704-727.

Wallerstein, I. 1974. *The Modern World System, Capitalist Agriculture and the Origins of the European World Economy in the Sixteenth Century*. New York: Academic Press.

Wooldridge, J. M. 2006. *Introductory Econometrics: A Modern Approach, Third Edition*. Cincinnati, OH: South Western College Publishing.

Zipf, G. K. 1946. “The P1 P2/D Hypothesis: On the Intercity Movement of Persons.” *American Sociological Review* 11(6):677-686.

Zlotnik, H. 1992. “Empirical Identification of International Migration Systems.” In *International Migration Systems: A Global Approach*. Edited by M.M. Kritz, L.L. Lim, and H. Zlotnik. Clarendon Press: Oxford.

Table 1: Covariates - Descriptions and Sources

	Description	Source
General		
Year	Continuous year term (0 = 2003)	
Geographic distance (*ij*)	Distance (km) between most populated cities of origin *i* and destination *j*	Centre D'Etudes Prospectives et D'Informations Internationales
Linguistic distance (*ij*)	One of more official languages shared by origin *i* and destination *j* (0 = no)	
Demographic		
Percent Urban (*i*,*j*)	Percent of population residing in urban areas at origin *i* and destination *j*	UN Population Division
Potential support ratio (*i*,*j*)	Ratio population ages 15-64 to 65+ multiplied by 100 at origin *i* and destination *j*	UN Population Division
Economic		
GDP per capita ratio (*j/i*)	GDP per capita at destination *j* is 1.05 times or more higher than at origin *i* (0 = no)	International Monetary Fund
Unemployment rate (*i*,*j*)	Rate of total unemployment at origin *i* and destination *j*	International Labor Organization
Change LF participation (*i*,*j*)	Percent change in rate of labor force participation from prior year at origin *i* and destination *j*	International Labor Organization
Geopolitical		
Social expenditures (*i*,*j*)	Total social expenditures on social protection per head at origin *i* and destination *j*	Eurostat
EU tenure (*ij*)	Relative tenure in the EU for origin *i* and destination *j*	Europa
	0 = neither *i* nor *j* were founding EU members	
	1 = *i* but not *j* was founding EU member	
	2 = not *i* but *j* was founding EU member	
	3 = both *i* and *j* were founding EU members	
Socio-cultural		
National/colonial origins (*ij*)	Origin *i* and destination *j* were ever same country or in colonial relationship (0 = no)	Centre D'Etudes Prospectives et D'Informations Internationales

Table 2: Descriptive Statistics

	2003	2004	2005	2006	2007	2003-07
Outcome:						
Harmonized bilateral migration flow (*ij*)	2767.92	1855.49	1960.23	2161.79	2523.85	2190.09
	(4422.40)	(6537.66)	(7015.25)	(7528.29)	(8867.26)	(7400.71)
						[0.89]
General:						
Geographic distance (*ij*)	1339.67	1424.95	1424.95	1424.95	1431.65	1420.08
	(728.06)	(768.59)	(768.59)	(768.59)	(746.14)	(759.57)
Linguistic distance (*ij*)	0.10	0.05	0.05	0.05	0.04	0.05
Demographic:						
Percent urban (*i,j*)	74.24	71.64	71.76	71.89	71.29	71.83
	(12.06)	(12.09)	(12.08)	(12.07)	(12.09)	(12.09)
						[0.98]
Potential support ratio (*i,j*)	4.33	4.65	4.65	4.47	4.46	4.54
	(0.56)	(0.68)	(0.68)	0.73	(0.71)	(0.698)
						[0.85]
Economic:						
GDP per capita ratio (*j/i*)	0.41	0.46	0.46	0.46	0.46	0.46
Unemployment rate (*i,j*)	6.53	8.25	8.35	7.99	7.21	7.81
	(2.62)	(3.89)	(3.70)	(3.19)	(2.42)	(3.32)
						[0.75]
Change LF participation (*i,j*)	48.46	43.62	43.46	43.19	42.71	43.63
	(12.67)	(11.46)	(11.97)	(12.61)	(12.84)	(12.37)
						[0.86]
Geopolitical:						
Social expenditures (*i,j*)	6758.71	4726.90	4840.36	4949.50	4713.49	4955.11
	(2319.57)	(3369.19)	(3442.87)	(3464.91)	(3585.63)	(3435.20)
						[0.95]
EU tenure (*ij*)						
0 = neither *i* nor *j* were founding EU members	0.34	0.57	0.57	0.57	0.60	0.56
1 = *i* but not *j* was founding EU member	0.26	0.19	0.19	0.19	0.18	0.19
2 = not *i* but *j* was founding EU member	0.26	0.19	0.19	0.19	0.18	0.19
3 = both *i* and *j* were founding EU members	0.14	0.05	0.05	0.05	0.04	0.06
Socio-cultural:						
National/colonial origins (*ij*)	0.05	0.07	0.07	0.07	0.06	0.06
N	210	600	600	600	702	2712

Mean and standard deviation for harmonized bilateral migration flow are for the year(s) listed in the columns.
Means and standard deviations for all covariates are for the prior year, *t-1*.
Standard deviations shown in parentheses and not shown for binary variables.
Proportion of variance explained *between* origin-destination pairs shown in brackets for time-varying covariates.

Table 3: Baseline OLS and FEVD Models of Bilateral Migration Flows Between EU Countries: 2003-2007

Outcome: Ln Harmonized bilateral migration flow (*ij*)	(1) OLS	(2) FEVD
Year	-0.002	0.073***
	(0.031)	(0.007)
GDP per capita ratio (*j/i*)	0.092	0.097**
	(0.087)	(0.041)
EU tenure (*ij*) = 1	1.210***	1.237***
	(0.109)	(0.022)
EU tenure (*ij*) = 2	1.355***	1.290***
	(0.108)	(0.045)
EU tenure (*ij*) = 3	2.609***	2.601***
	(0.177)	(0.035)
National/colonial origins (*ij*)	1.624***	1.633***
	(0.162)	(0.030)
Residual unit fixed effect, *h*		0.999***
		(0.002)
Constant	4.864***	4.697***
	(0.101)	(0.022)
Observations	2712	2010
Root Mean Square Error	2.06	0.277

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: FEVD Models of Bilateral Migration Flows Between EU Countries: 2003-2007

Outcome: Ln Harmonized bilateral migration flow (*ij*)	(3) FEVD	(4) FEVD	(5) FEVD
Year	0.069***	0.052***	0.019***
	(0.007)	(0.008)	(0.006)
Ln Geographic distance (*ij*)	-1.053***	-1.083***	-0.794***
	(0.006)	(0.006)	(0.007)
Linguistic distance (*ij*)	0.634***	0.916***	0.136***
	(0.026)	(0.022)	(0.012)
Ln Percent Urban (*i*)	0.569***	0.491***	0.246***
	(0.023)	(0.029)	(0.044)
Ln Percent Urban (*j*)	1.002***	0.397***	0.047
	(0.028)	(0.035)	(0.035)
Ln Potential support ratio (*i*)	-3.148***	-3.172***	-2.003***
	(0.020)	(0.023)	(0.057)
Ln Potential support ratio (*j*)	-3.148***	-3.221***	-1.800***
	(0.016)	(0.016)	(0.067)
GDP per capita ratio (*j/i*)		0.232***	0.102***
		(0.017)	(0.027)
Ln Unemployment rate (*i*)		1.221***	1.739***
		(0.021)	(0.011)
Ln Unemployment rate (*j*)		1.363***	1.931***
		(0.051)	(0.034)
Ln Change LF participation (*i*)		1.801***	1.238***
		(0.035)	(0.053)
Ln Change LF participation (*j*)		2.390***	1.696***
		(0.049)	(0.069)
Ln Social expenditures (*i*)			0.450***
			(0.025)
Ln Social expenditures (*j*)			0.658***
			(0.027)
EU tenure (*ij*) = 1			0.516***
			(0.022)
EU tenure (*ij*) = 2			0.404***
			(0.034)
EU tenure (*ij*) = 3			0.810***
			(0.033)
National/colonial origins (*ij*)			1.028***
			(0.032)
Residual unit fixed effect, *h*	0.999***	0.998***	0.998***
	(0.001)	(0.006)	(0.003)
Constant	15.632***	-1.958***	-11.996***
	(0.113)	(0.324)	(0.115)
Observations	2010	2010	2010
Root Mean Square Error	0.277	0.275	0.274

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

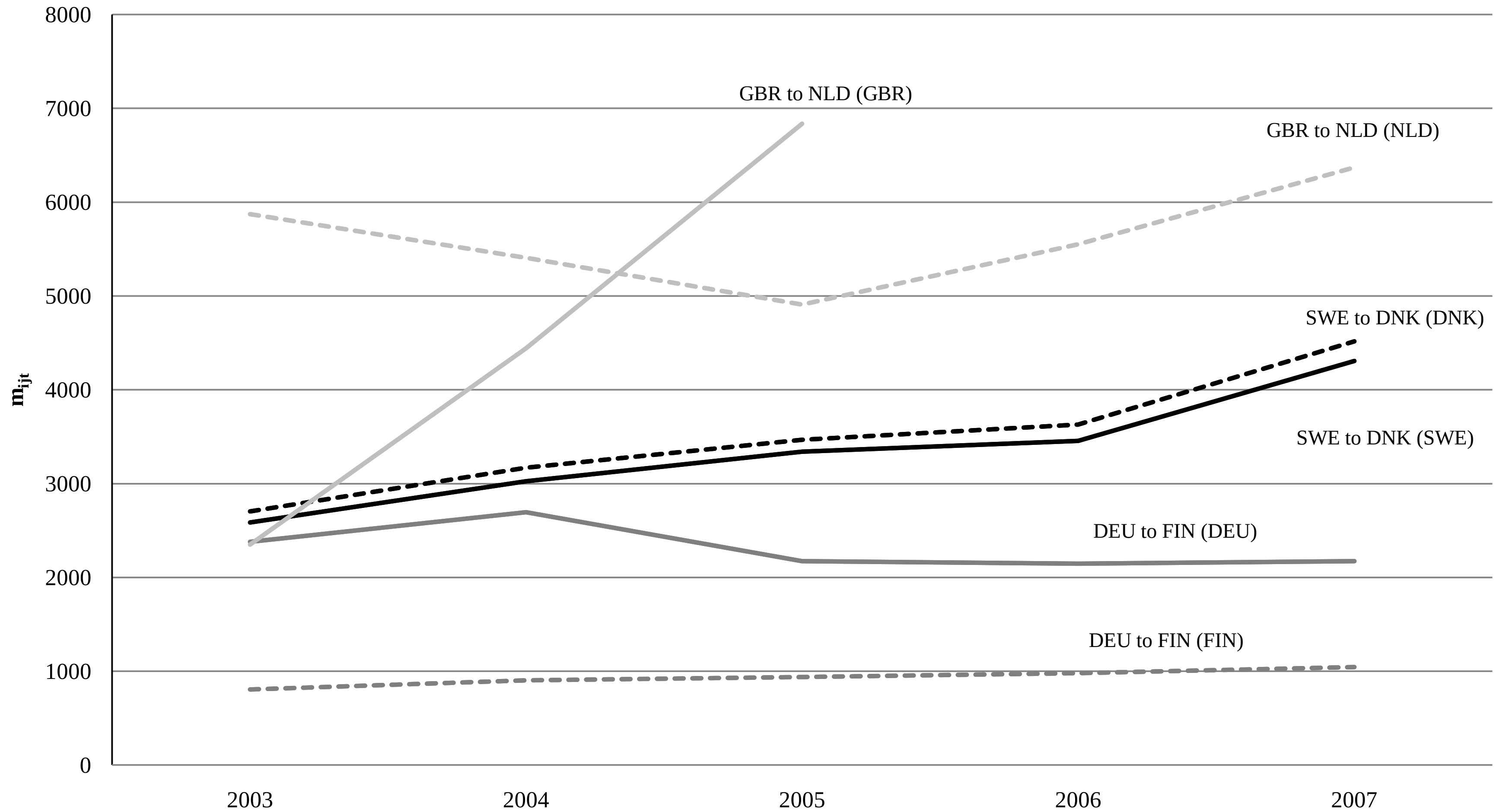
Figure 1: Double Count Matrix of Bilateral Migration Flows Between EU-15 Countries: 2003

		AUT	BEL	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	LUX	NLD	PRT	SWE
	E		177	4422	100	300	196	426	668	340	68	852	44	295	180	396
AUT	*I*		…	13456	262	554	92	…	…	…	…	…	7	510	33	333
	E/I		…	0.33	0.38	0.54	2.13	…	…	…	…	…	6.29	0.58	5.45	1.19
	E	…		…	…	…	…	…	…	…	…	…	…	…	…	…
BEL	*I*	267		4291	587	3037	184	…	…	…	…	…	804	5348	105	399
	E/I	…		…	…	…	…	…	…	…	…	…	…	…	…	…
	E	15976	4623		2712	16236	2380	19060	15550	18106	2415	33802	1510	8616	8880	3786
DEU	*I*	12239	…		3221	13746	807	…	…	…	…	…	436	7921	645	2872
	E/I	1.31	…		0.84	1.18	2.95	…	…	…	…	…	3.46	1.09	13.77	1.32
	E	231	511	2540		1720	403	1333	4317	229	264	782	131	609	174	3786
DNK	*I*	203	…	2693		764	371	…	…	…	…	…	14	474	58	2872
	E/I	1.14	…	0.94		2.25	1.09	…	…	…	…	…	9.36	1.28	3.00	1.32
	E	93	647	2109	130		102	2474	2335	38	487	801	89	600	627	164
ESP	*I*	615	…	14647	1665		68	…	…	…	…	…	28	2794	473	1234
	E/I	0.15	…	0.14	0.08		1.50	…	…	…	…	…	3.18	0.21	1.33	0.13
	E	76	245	761	397	792		284	1070	56	110	210	57	217	26	3428
FIN	*I*	251	…	2204	421	802		…	…	…	…	…	2	362	15	3395
	E/I	0.30	…	0.35	0.94	0.99		…	…	…	…	…	28.50	0.60	1.73	1.01
	E	…	…	…	…	…	…		…	…	…	…	…	…	…	…
FRA	*I*	741	…	18133	1488	8847	312		…	…	…	…	987	2919	458	931
	E/I	…	…	…	…	…	…		…	…	…	…	…	…	…	…
	E	…	…	…	…	…	…	…		…	…	…	…	…	…	…
GBR	*I*	1180	…	13197	3707	34117	914	…		…	…	…	37	5872	947	3022
	E/I	…	…	…	…	…	…	…		…	…	…	…	…	…	…
	E	…	…	…	…	…	…	…	…		…	…	…	…	…	…
GRC	*I*	465	…	12959	278	273	63	…	…		…	…	5	882	15	585
	E/I	…	…	…	…	…	…	…	…		…	…	…	…	…	…
	E	…	…	…	…	…	…	…	…	…		…	…	…	…	…
IRL	*I*	138	…	2046	306	1649	146	…	0	…		…	3	615	37	230
	E/I	…	…	…	…	…	…	…	…	…		…	…	…	…	…
	E	…	…	…	…	…	…	…	…	…	…		…	…	…	…
ITA	*I*	1460	…	23702	895	5796	209	…	…	…	…		68	1661	312	473
	E/I	…	…	…	…	…	…	…	…	…	…		…	…	…	…
	E	22	1119	747	119	73	33	1254	171	22	44	208		97	521	74
LUX	*I*	57	…	1728	196	89	34	…	…	…	…	…		166	10	78
	E/I	0.39	…	0.43	0.61	0.82	0.97	…	…	…	…	…		0.58	52.10	0.95
	E	470	9284	9822	430	3365	292	3373	7022	482	459	1274	150		666	648
NLD	*I*	655	…	13015	820	3567	239	…	…	…	…	…	25		264	707
	E/I	0.72	…	0.75	0.52	0.94	1.22	…	…	…	…	…	6.00		2.52	0.92
	E	0	0	955	0	0	0	849	2187	0	0	0	770	0		0
PRT	*I*	330	…	7699	170	5505	56	…	…	…	…	…	512	1619		143
	E/I	0.00	…	0.12	0.00	0.00	0.00	…	…	…	…	…	1.50	0.00		0.00
	E	238	411	1580	2585	1356	3386	946	3676	510	205	441	66	499	92	
SWE	*I*	474	…	3397	2705	1537	3438	…	…	…	…	…	11	638	31	
	E/I	0.50	…	0.47	0.96	0.88	0.98	…	…	…	…	…	6.00	0.78	2.97	

E - Emigration report of origiin; *I* - Immigration report of destination; *IE* - Emigration/Immigration report

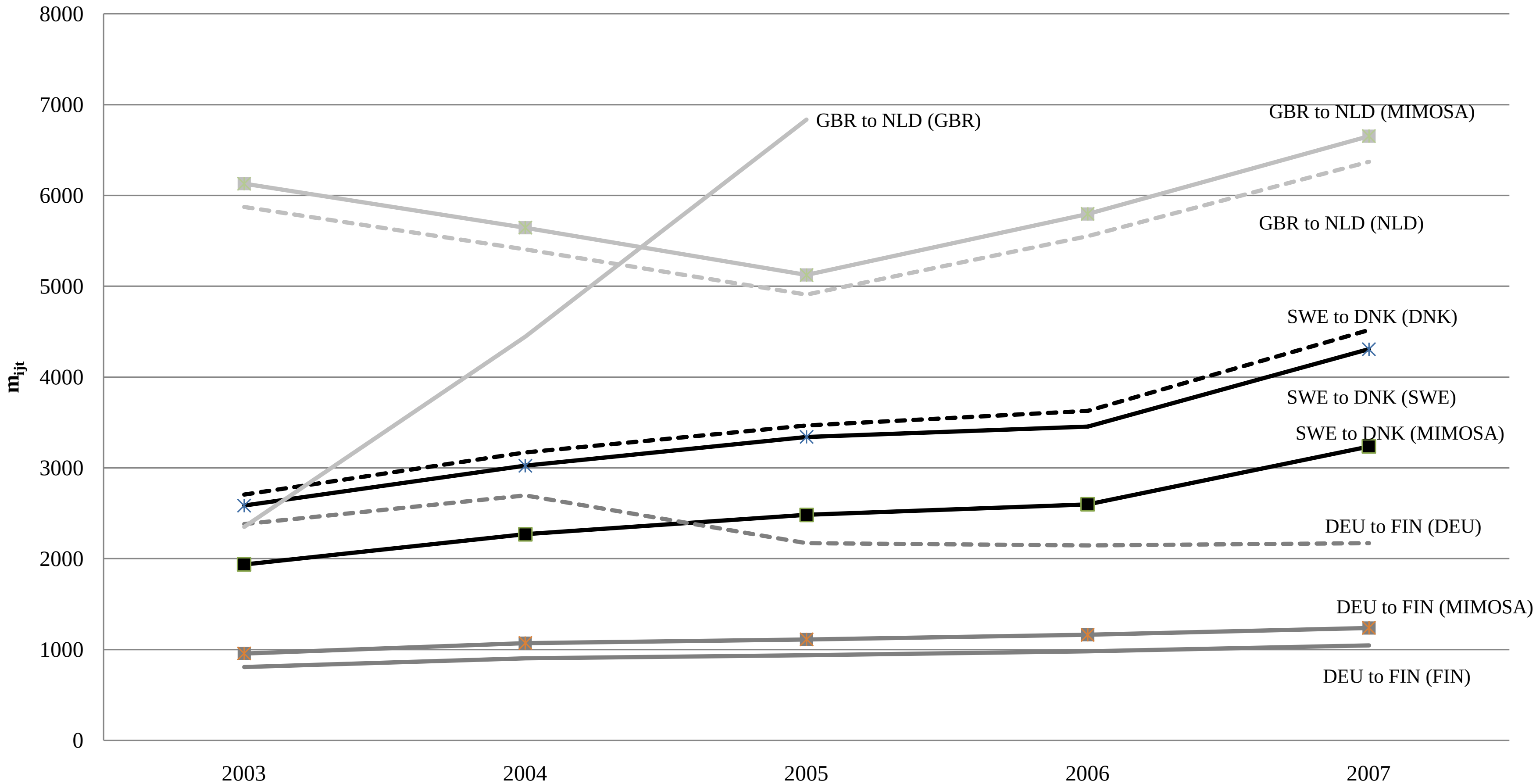
Source: Kupiszewska & Nowak (2008)

Figure 2: Bilateral Migration Flows Between Selected EU-15 Countries: 2003-2007
(Reporting country shown in parentheses)



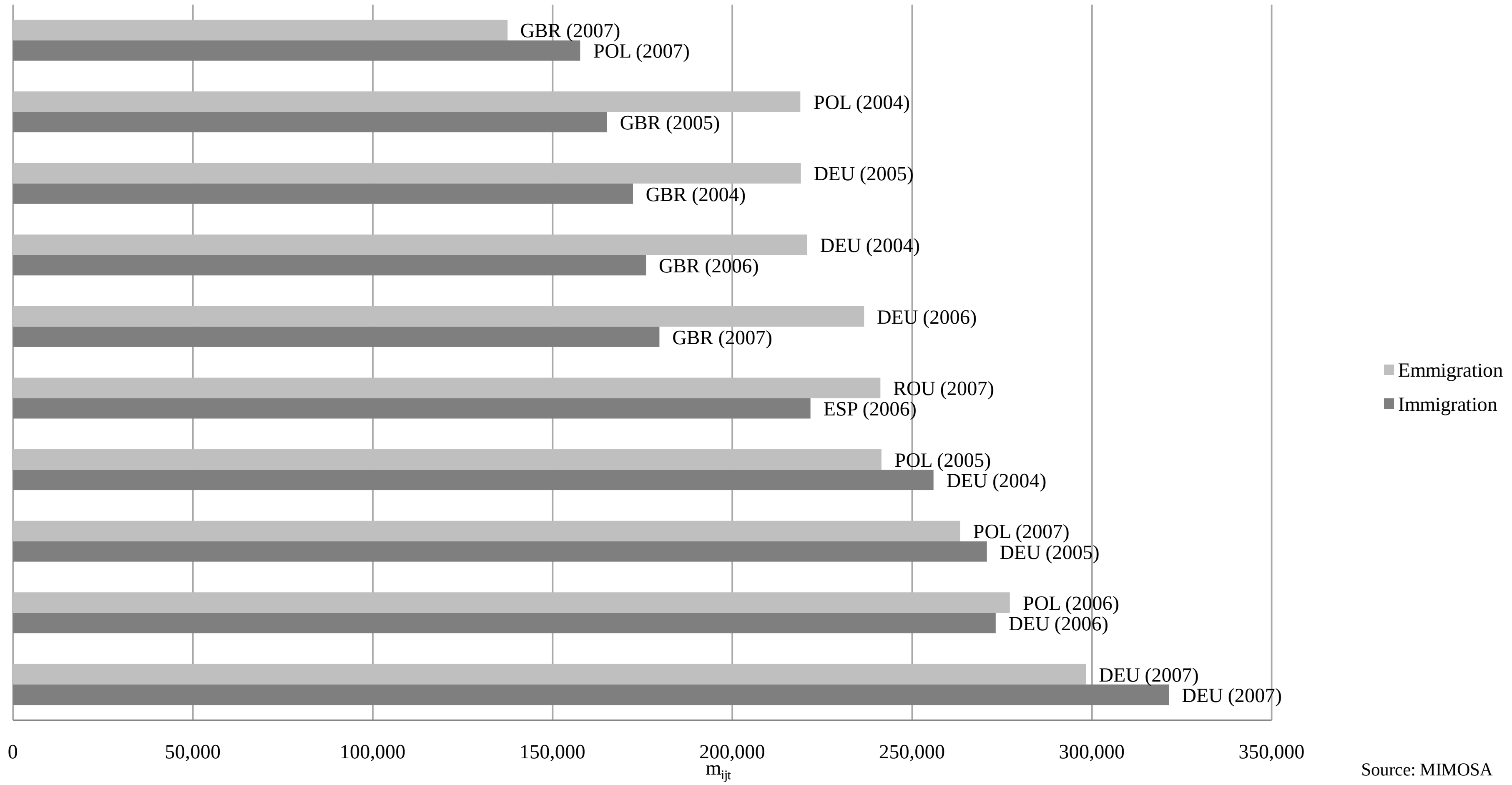
Source: Eurostat

Figure 3: Harmonized Bilateral Migration Flows Between Selected EU-15 Countries: 2003-2007
(Reporting country or MIMOSA shown in parentheses)



Source: Eurostat

Figure 4: Top 10 Emmigration and Immigration Country-Years



Source: MIMOSA

Figure 5: Ratio of Predicted to Baseline Migration Flows