Multilateral Attractiveness, Migration Networks, and Destination Choices of International Migrants to the Madrid Metropolitan Area

Ludo Peeters

Department of Business Economics, Hasselt University,
Agoralaan, Gebouw D, 3590 Diepenbeek (Belgium)
Email: ludo.peeters@uhasselt.be; Tel.: +32 11 268635; Fax: +32 11 268700

Coro Chasco*

Department of Applied Economics, Autonomous University of Madrid,
Avda. Francisco Tomás y Valiente 5, 28049 Madrid (Spain)
Email: coro.chasco@uam.es; Tel: + 34 91 474266; Fax: + 34 91 4973943

[Version January 2013]

ABSTRACT

We investigate the effects of local characteristics on destination choices of international migrants to the Madrid metropolitan area. It is argued that the choice of a particular location does not only depend on the attractiveness of that location but also on the attractiveness of all other locations in the choice set. This is likely to be more important with a narrowly defined choice set with a fairly high degree of substitutability between functionally similar locations. We use the term "multilateral attractiveness" to refer to the unobserved influence exerted by all locations in the spatial choice set. Taking advantage of the equivalence relation that exists between Conditional Logit and Poisson, we estimate a location-choice model using the Poisson pseudo maximum likelihood estimator. Location-specific effects have been incorporated to account for unobserved spatial similarities (i.e., for possible violations of the IIA assumption). The fixed-effects Poisson model is applied to municipality-level migration data, for a set of five broadly defined groups of immigrants. The proposed estimation strategy has important empirical implications, where the magnitudes and/or signs of the estimated coefficients change in the expected direction. We find that the spatial pattern of immigrants' location choices is fairly persistent over the time span considered. We also find that the impacts of local ethnic communities (network effects) are insignificant in some instances, while suggesting hetero-local settlement preferences or possible in-group rivalry in other instances.

Keywords:

International migration
Multilateral attractiveness
Spatial structure
Migration networks
Hetero-local preferences/in-group rivalry

1 Introduction

Over the last two decades, Spain has experienced an unprecedented increase in immigration. The most important destination of immigrants entering Spain is the Madrid metropolitan area (MMA), which received 157,000 new immigrants in 2009 (or about 12% of the total number of new immigrants to Spain in that year). The large influx of foreigners has become an issue of public concern in Spain, as the employment situation of many immigrants (particularly those coming from developing countries) has deteriorated sharply with the economic crisis that started in 2007/08.

The present paper focuses on the location choices of international immigrants to the MMA, using municipality-level migration data. As the empirical literature on immigrants' location choices at this fine spatial level is rather scarce, this study intends to fill some of the gaps. Specifically, our purpose is to provide some insightful information on which municipalities within the MMA can expect larger (or smaller) numbers of new immigrants. This is an important issue from a policy point of view, which can be imperative for, say, the need to provide public (social) services in certain locations.

To obtain a better understanding of the local determinants of immigrants' location choices, we present a modeling strategy that aims to account for what we term *multilateral attractiveness*. The basic intuition behind the notion of multilateral attractiveness is the following. If multiple locations within a given choice set are fairly close substitutes for one another—which is most likely to be the case with multiple choice alternatives in a relatively small geographical area (given their potentially high degree of functionally similarity), the number of immigrant arrivals in a given location is not only dependent on the (unilateral) attractiveness of that location but rather on the *relative* (multilateral) attractiveness of *all* locations contained in the choice set. As a result, one should not only look at the local attributes (e.g., local employment opportunities), but also at the contextual attributes (i.e., the attributes of the surrounding locations). Moreover, this contextual dependence is likely to vary with the relative spatial position of a location within the metropolitan area. This paper, therefore, presents a *context-sensitive* model to identify local determinants of immigrants' destination choices that, besides capturing the influences of the wider (supra-local) choice environment, also recognizes the *spatial structure* of the choice alternatives.[1]

Location-choice models that ignore the multilateral nature of a location's attractiveness typically produce estimates that are biased and inconsistent due to the presence of omitted variables (see, e.g.,

[1] Specific interest in the multilateral influences on location decisions has only recently made its entrance in empirical studies of international migration. Mayda (2010) was among the first to adopt the idea of multilateral attractiveness (as used here) in the context of international migration, which she termed *multilateral pull effects*, yet using an "atheoretical" (ad-hoc) measure (the average per worker GDP levels over all the destination countries, each weighted by the inverse of their physical distance from the origin country). In contrast, Bertoli and Fernández-Huertas Moraga (2013) introduced the notion of *multilateral resistance to migration*, and estimated a theoretically-founded Nested-Logit type model of international migration. Notice also the close affinity with the notion of *multilateral resistance to trade* introduced by Anderson and van Wincoop (2003).

Hanson, 2010). Specifically, parameter estimates will generally be *smaller* (understated) if local factors are confounded with relative attractiveness. In fact, variations in *multilateral* attractiveness tend to re-direct immigrants across locations but may imply only small changes in *relative* attractiveness of any given location, and consequently so in the number of migrant arrivals in each location; that is, certain locations gain some immigrants, while others loose some. Therefore, if multilateral influences are not accounted for, the small shifts in the distribution of immigrants across locations will be erroneously ascribed to *local* attributes, so that the estimated local effects have no causal interpretation.

Given the discrete nature of migrants' destination choices, the latter are usually modeled within a Conditional Logit (CL) framework (e.g., Davies et al., 2001; Scott et al., 2005; Åslund, 2005; Brown and Scott, 2012). The appeal of CL lies in its consistency with the random utility maximization (RUM) framework (McFadden, 1974). However, this paper estimates a model of location choices using the Poisson pseudo maximum likelihood (PPML) estimator, thereby taking advantage of the equivalence relation between CL and Poisson. Our strategy therefore builds on the findings of Guimarães et al. (2003, 2004) and Schmidheiny and Brülhart (2011), who have shown that PPML returns parameter estimates that are identical to those implied by CL with grouped data and *group-specific effects*, where we use the immigrants' origin countries (or sets of countries) as the grouping variable.

One important problem still remains, however. This problem is related to the Independence of Irrelevant Alternatives (IIA) assumption underlying CL (following directly from the assumption that the errors are independent and identically distributed), which is naturally carried over to Poisson.[2] Specifically, IIA means that immigrants look at all locations as similar (substitutes for one another), conditional on the observable local attributes. This is a too strong assumption, though, as the choice alternatives are unevenly distributed over space. That is, locations situated in close proximity are more likely to be substitutes for one another than are locations situated at greater distances from each other, hence potentially giving rise to spatially correlated location choices. Moreover, the spatial pattern of location choices can be quite persistent over time. In view of these problems, we incorporate *location fixed effects* to account for the spatial arrangement of the choice alternatives. Although fixed effects are primarily designed to deal with cross-sectional heterogeneity due to unobservable location-specific attributes, it is anticipated that they are inherently spatial and, therefore, capable of capturing unobserved *spatial similarities* (spatially clustered omitted variables). Of course, whether the location-specific effects will be sufficient to ensure cross-sectional error independence ultimately remains an empirical question, but this can easily be tested (as we will do in the empirical part of the paper).

[2] In the spatial analysis literature, the IIA assumption has occasionally been called the *Independence of Spatial Structure* (ISS) assumption (e.g., Borgers and Timmermans, 1988), and implies that the probability of choosing an alternative is independent of the spatial configuration of all the alternatives in the choice set, so that any spatial correlation is simply assumed away.

We test the implications of multilateral attractiveness for location choices of five (broadly defined) groups of immigrants to the MMA, originating from the EU25 countries (excluding Spanish nationals), Bulgaria–Romania, Latin America, Morocco, and China. Unlike most other studies of international migration, the empirical analysis is conducted at a finely grained spatial scale, using the *municipalities* within the MMA as the spatial unit.[3] We assume that the decision to settle *somewhere* in the MMA *precedes* the choice of any given municipality. Moreover, all immigrants are assumed to have already passed the "cliff at the border" (see Bertoli and Fernández-Huertas Moraga, 2012); all immigrants have either resolved their visa issues, etc., or have simply entered Spain undocumented. Consequently, all immigrants are considered *footloose*, as (after having entered Spain) they are free to choose whichever place to reside. This, in turn, also means that the present study is closely allied to studies of industrial location (e.g., Guimarães et al., 2004; Arauzo-Carod et al, 2010; Brülhart et al., 2012) and competing destinations in the context of internal migration (e.g., Pellegrini and Fotheringham, 2002), involving complex choice scenarios where decision makers confront many narrowly defined spatial alternatives.

A final concern addressed in this paper relates to the measurement of the impact of pre-existing *migrant stocks* (i.e., immigrant communities already in place at a location) on new immigrants' location choices. There is ample empirical evidence suggesting that immigrants are spatially clustered (e.g., Edin et al., 2003; among others); that is, newly arriving immigrants tend to settle in *ethnic enclaves* because of the value of nearby ethnicity-related amenities and positive social network externalities (reducing the costs and/or risks of migration by allowing newly arriving immigrants to gain easier access to jobs and facilitate adaptation to the new environment). However, with multiple potential locations in a narrowly defined spatial choice set it is not obvious that we should see a *local* correlation between the number of migrant arrivals and the size of the pre-existing migrant stock. There are at least two (interrelated) reasons why migrant-stock effects may turn out to be small, negligible, or even negative. A first reason is that new immigrants may choose to settle in widely dispersed locations rather than in concentrated ethnic communities, while still maintaining their ethnic identity. This phenomenon, which Zelinsky and Lee (1998) termed *hetero-localism*, implies that *positive* network externalities may stretch out far beyond the local level.[4] A second reason is that *negative* externalities

---

[3] Most earlier studies of international migration are typically conducted at the *country* level, covering either multiple sending and one receiving country (e.g., Clark et al., 2007; Hatton and Williamson, 2011; Bertoli and Fernández-Huertas Moraga, 2013), one sending country and multiple receiving countries (McKenzie et al., 2012), or multiple sending and receiving countries (e.g., Pedersen et al., 2008; Ortega and Peri, 2009; Mayda, 2010; Bertoli and Fernández-Huertas Moraga, 2012). Other studies would focus on immigrants' settlement in different *metropolitan areas* (MSA's) or *regions* (states) within a country, mainly in the U.S. (e.g., Newbold, 1999; Zavodny, 1999; Dodson, 2001; Munshi, 2003; Kaushal, 2005; Jaeger, 2008).

[4] The notion of hetero-localism, introduced by Zelinsky and Lee (1998), maintains that, even though immigrants may settle in widely dispersed locations, ethnic group members can stay closely "connected" (through recent advances in information and communication technology, improved transportation facilities, etc.). In the presence

(or diseconomies of size) may arise if immigration is subject to *adverse selection* (for example, in the case of the large number of undocumented Romanians arriving in the MMA), or when growing ethnic concentrations intensify within-group *employment competition* among similarly-/low-skilled workers, hence exerting a downward pressure on wages in those labor-market segments they gravitate toward (Bauer et al., 2007) or inducing local crowding out (Liu, 2013). In other words, a trade-off could be at work, whereby newly arriving immigrants weigh the advantages associated with the ethnic community against possible disadvantages (Gang and Rivera-Batiz, 1994).[5] Therefore, if (for whatever reason) the latter are perceived to outweigh the former, new immigrants may want to disperse their settlement to non-traditional destinations.[6]

The rest of the paper is organized as follows. Section 2 provides some background information on basic characteristics of the MMA, recent immigration to the MMA, and the spatial settlement patterns of newly arriving immigrants from the five origin-groups considered. Section 3 sets out on the setup of the model of location choices and the fixed-effects Poisson estimator used. Section 4 presents the empirical model specification, followed by a discussion of the econometric results as well as those from some diagnostic tests. Section 5 provides a summary and some concluding remarks.

2 Background: some basic empirics

In this section, we provide some background information that could be helpful in appreciating the empirical results presented later in the paper. First, we briefly discuss a number of basic characteristics of the MMA, which is our study area. Second, we provide some facts and figures about recent immigration to the MMA. Finally, we take a somewhat closer look at the spatial settlement patterns of new immigrants to the MMA.

2.1 Basic characteristics of the MMA

The MMA includes the central city of Madrid and 40 surrounding municipalities (see the base map in Appendix C, Fig. C.1). The MMA has a population of about 5.8 million people (of which about 3.2 million, or 55%, in the city of Madrid), and covers an area of 2,700 square kilometers. So, the geographical size of the study area is relatively small.

The spatial structure of the MMA is coherent with an historical mono-centric structure, which has been built upon the interdependence between the central city (Inner Madrid) and the suburban

---

of hetero-local settlement, both dispersion and clustering can occur at the same time, hence potentially giving rise to non-contiguous nodes of (moderate) ethnic concentrations.

[5] An interesting case study of in-group rivalry within the Polish community in Brussels can be found in Grzymala-Kazlowska (2005).

[6] A third reason may be the within-group heterogeneity (composition) of the broadly defined immigrant groups, particularly the EU25 and Latin-American immigrant groups.

reaches, generating a series of metro rings. Yet, as in many other metropolitan areas in the developed world, a process of suburbanization took place, which was driven by several factors, including greater availability of land in the ring belt as compared to the urban core, the de-concentration of economic activities, lower housing prices in the suburban fringe, and a dense network of public transportation facilities and radial highways, giving rise to the emergence of "edge cities" (i.e., a structure of multi-centricity).

2.2 Recent immigration to the MMA

The MMA provides an interesting case study. The metropolitan area is an emerging hub in the global economy, which has reached a high level of international competitiveness during the last two decades. The concentration of national investments in the MMA has played a key role in promoting Madrid's international accessibility (OECD, 2007). As a consequence, the MMA has been the largest recipient of immigrants in Spain, with the booming economy being a significant factor in explaining the significant growth in the number of migrant arrivals.[7] Immigrants to the MMA area come from all over the world. In 2009, 84,323 (49.4%) immigrants come from America, 45,825 (26.9%) from Europe, 23,929 (14.0%) from Africa, 16,406 (9.6%) from Asia, and only 76 (0.04%) from Oceania. However, in the present paper, we focus on five broadly defined origin-groups of immigrants: EU25 (excluding Spanish nationals), Bulgaria–Romania, Latin America (Spanish speaking), Morocco, and China.

There are three reasons for selecting these immigrant groups. First, these groups together represent about 85% of the total number of immigrants to the MMA in the period 2005–2009, and about 90% of the total number of immigrants to the whole region of Madrid (see Table 1). Second, these groups provide cases of different ethnic and religious identities as well as different linguistic backgrounds, educational levels, professional skills, etc. However, the selected groups are likely to be heterogeneous in terms of socio-economic characteristics, given the different nationalities of the immigrants within each of the broadly defined groups. Third, the groups have different settlement histories: Latin Americans and Moroccans (as well as EU25 immigrants) have already a relatively long immigration history in Spain, whereas the large influx of Bulgarian–Romanian immigrants and (in particular) Chinese immigrants is of a more recent date (see, e.g., Mallet, 2011).

< Insert **Table 1** about here>

An exploratory shift-share analysis (see Appendix C, Table C.1) reveals that the city of Madrid

[7] The economic crisis in Spain, which started in the summer of 2007, would put an end to the immigration boom, albeit with some time lag, as substantial inflows persisted until the last part of 2008. In 2009, a decline in migrant inflows started to set in.

looses immigrants in favor of the rest of the metro area (holding constant the composition), whereas it gains immigrants at the expense of the rest of the metro area due to shifts in the composition of the total flow of immigrants to the MMA (e.g., more EU25 immigrants in 2009). However, the "group"-gains are insufficient to offset, or outweigh, the "area"-losses.

2.3 Spatial settlement pattern of immigrants to the MMA

The spatial settlement patterns of new immigrants to the MMA do not appear to conform to the quintessential image of concentrations in high-density, low-quality, inner-city locations. Upon their arrival, immigrants are scattered throughout the suburban reaches of the MMA, with only a moderate degree of spatial clustering in certain parts that hardly qualify as classic "ethnic ghettos".

Fig. 1 presents a set of maps showing the settlement patterns for the five selected groups of immigrants. These maps are constructed on the basis of the *location quotients* for the 2009 immigration rates. The location quotient (LQ) can be thought of as sort of specialization index, which is defined as $LQ_{ij} = (n_{ij}/P_j)/(n_i/P)$, where n_{ij} is the number of migrant arrivals from group i in location j, P_j is the total population in location j, n_i is the total number of migrant arrivals from group i to the MMA, and P is the total population of the MMA. If $LQ_{ij} = 1$, location j has the same percentage of immigrants from group i as the MMA as a whole; if $LQ_{ij} > 1$, group i is over-represented in location j; and if $LQ_{ij} < 1$, group i is under-represented in location j.

<Insert **Fig. 1** about here>

The maps in Fig. 1 reveal that newly arriving immigrants settle in multiple locations throughout the MMA. Residential dispersion varies considerably from one immigrant group to another. Certainly, some concentrated areas can be discerned, but this is far from an overarching pattern, with noticeable differences across immigrant groups, roughly going from highly a dispersed settlement (Latin-Americans) over a moderately dispersed settlement with some areas of concentration (Bulgarians–Romanians, Moroccans, and EU25 immigrants) to a concentrated settlement (Chinese immigrants). Thus, the overall picture is one of variation in what appears to be a spatially dispersed settlement pattern, reflecting some kind of *nodal* hetero-localism (Hardwick, 2006).

What is strikingly visible from the maps in Fig. 1 is that immigrants tend to "bypass" the central city of Madrid, relatively speaking, and to settle in suburban areas.[8] The suburbanization propensity of immigrants could be explained by metropolitan de-concentration; that is, the changing distribution of job opportunities within the MMA and the emergence of "edge cities," which are characterized by

[8] Note, however, that about one-third to almost 60% of new immigrants still arrive in the city of Madrid (see panel B of Table 1), which is thus mainly a matter of scale.

an increasing share of the metro area's employment and the concomitant settlement pattern of new migrants following spatially dispersed employment opportunities (Newbold and Spindler, 2001).[9]

3 Model setup and estimation

In this section, we begin with explaining the equivalence relation between Conditional Logit (CL) and Poisson. Next, we discuss the fixed-effects Poisson model to cope with the issue of cross-sectional (or spatial) error correlations.

3.1 Equivalence relation between Conditional Logit and Poisson

Consider immigrants from group i ($i = 1, \dots, I$), who independently select a destination j from a set of J potential destinations ($j = 1, \dots, J$). It is assumed that the indirect utility of an individual migrant m from group i at destination j can be approximated by the following linear *random-utility model* (RUM), including a stochastic term (see also Davies et al., 2001, p. 340):

$$U_{mj|i} = \boldsymbol{\beta}'\mathbf{y}_{j|i} + \varepsilon_{mj|i} \tag{1}$$

where $\mathbf{y}_{j|i}$ is a ($J \times 1$) vector of destination-specific characteristics that all immigrants from group i have before them, and $\boldsymbol{\beta}$ is a ($K \times 1$) vector of unknown parameters to be estimated. Then, the CL model is defined by assuming that the stochastic term $\varepsilon_{mj|i}$ is independent across m and j and follows an Extreme Value type-1 distribution (reflecting the idiosyncrasies specific to each individual as well as the unobserved attributes of the destinations in the choice set). The probability that an individual migrant m chooses destination j rather than another destination $k \neq j$ is

$$P_{mj|i} = P_{j|i} = \frac{e^{\boldsymbol{\beta}'\mathbf{y}_{j|i}}}{\sum_{k=1}^{J} e^{\boldsymbol{\beta}'\mathbf{y}_{k|i}}} \tag{2}$$

where $\sum_j P_{mj|i} = 1$, for all m and each i.

Eq. (2) assumes that all individuals are affected *symmetrically* by the local attributes contained in vector $\mathbf{y}_{j|i}$ (hence, $P_{mj|i} = P_{j|i}$), from which it follows that all immigrants belonging to the *same* group i have identical preferences and derive equal utility from choosing destination j.[10] It further means that $P_{j|i}$ represents the *proportion* of immigrants from group i that chooses destination j. On the other hand, immigrants' preferences can be (and likely will be) different *across* groups.

---

[9] Examination of the factors involved in shaping the hetero-local settlement patterns is outside the scope of the present paper. Interested readers are directed to other (sociological) studies on hetero-localism; e.g., Newbold and Spindler (2001), Hardwick and Meacham (2005), and Hardwick (2006).

[10] As a result, differences in unobserved characteristics of individual immigrants may be a source of concern (e.g., Grogger and Hanson, 2011). However, examination of this issue is outside the scope of the present paper.

The CL model implicitly assumes that the total number of immigrants from origin i to the metro area, $n_i = \sum_{j=1}^{J} n_{ij}$, is fixed and does not depend on the location-specific attributes (Schmidheiny and Brülhart, 2011, p. 215). Then, the *expected number* of immigrants from group i choosing destination j is

$$E(n_{ij}) = n_i P_{j|i} = n_i \frac{e^{\boldsymbol{\beta}'\mathbf{y}_{j|i}}}{\sum_{k=1}^{J} e^{\boldsymbol{\beta}'\mathbf{y}_{k|i}}} \tag{3}$$

and the *stochastic* version of Eq. (3) is

$$n_{ij} = n_i \frac{e^{\boldsymbol{\beta}'\mathbf{y}_{j|i}}}{\sum_{k=1}^{J} e^{\boldsymbol{\beta}'\mathbf{y}_{k|i}}} u_{ij} \tag{4}$$

Eq. (4) can now be written in *multiplicative form* as

$$n_{ij} = e^{\alpha_i + \boldsymbol{\beta}'\mathbf{y}_{j|i}} u_{ij} \tag{5}$$

where $\alpha_i = \ln n_i - \ln \sum_{k=1}^{J} e^{\boldsymbol{\beta}'\mathbf{y}_{k|i}}$ is a group-specific effect, and $IV_{\bullet|i} = \ln \sum_{k=1}^{J} e^{\boldsymbol{\beta}'\mathbf{y}_{k|i}}$, called the *inclusive value* (within a Nested Logit framework), represents the expected utility that immigrants obtain from *all* destinations in the choice set (e.g., Pellegrini and Fotheringham, 2002). Interestingly, the model in Eq. (5) can act as a group-level regression equation which can be estimated using the Poisson pseudo maximum likelihood (PPML) estimator, following Guimarães et al. (2004).

The important point here is that the α_i term in Eq. (5) neatly aligns with the notion of multilateral attractiveness, where α_i controls for the fact that immigrants' choices of a destination always depend on the utility attached to *all* destinations in the choice set. Therefore, if multilateral attractiveness is not accounted for, the unobserved α_i ends up in the error term, v_{ij}, hence giving rise to an *endogeneity* problem. Given the potentially high degree of substitutability that exists between the destinations in a narrowly defined spatial choice set, the omitted-variable biases may be quite severe (McFadden, 1984, p. 1422).

3.2 Estimation strategy: introducing location-specific effects

While the model in Eq. (5) serves as a starting point for the empirical implementation of the model, where the inclusion of group-specific effects α_i mitigates the problem associated with omitted factors that immigrants may consider important when deciding where to settle, it does not provide a basis for taking into account the *spatial patterns* that may arise from their utility-maximizing location choices. When dealing with data at the level of municipalities in a narrowly defined choice set, both observed and unobserved local factors are likely to extend their influence beyond the boundaries of the spatial units (i.e., spatial externalities). Therefore, Eq. (5) almost certainly fails to ensure *cross-sectional error* *independence*—or the IIA assumption to hold. Because the group-specific effects α_i only allow for unobserved heterogeneity across groups of immigrants (as in Ortega and Peri, 2009; Cadena, 2013), they are *invariant* across *locations* and, therefore, unlikely to guarantee i.i.d. errors.

To account for the spatial context of the local factors affecting immigrants' destination choices, we embed the Poisson model into a *two-period* framework, hence providing us with an additional time dimension, and incorporate *location-specific effects* (Guimarães et al., 2004). Specifically, we estimate a conditional *fixed-effects* Poisson model aimed at capturing unobserved location attributes that may have a similar value to all immigrants within the same group but may have a dissimilar value to immigrants from distinct groups (see also, e.g., Cadena, 2013). This can be achieved by reformulating the model as

$$n_{ijt} = e^{\gamma_{ij}+\alpha_{it}+\boldsymbol{\beta}_1'\mathbf{y}_{jt}+\boldsymbol{\beta}_2'\mathbf{x}_{ijt}}\, u_{ijt} \tag{6}$$

where $\mathbf{x}_{ij}$ is an additional (J × 1) vector of origin-specific local characteristics (e.g., migrant stocks), and $\boldsymbol{\beta}_2$ is an additional (G × 1) vector of unknown parameters to be estimated, α_{it} are *group-time effects*, and γ_{ij} are *group-location fixed effects*. Because $\alpha_{it} = \ln n_{it} - \ln \sum_{k=1}^{J} e^{\gamma_{ij}+\boldsymbol{\beta}_1'\mathbf{y}_{jt}+\boldsymbol{\beta}_2'\mathbf{x}_{ijt}}$ in Eq. (6), it follows that the group-time effects implicitly comprise the group-location-specific effects γ_{ij} for all j.[11]

If the model in Eq. (6) is capable of accommodating the correlations that exist among unobservable localized factors across destinations, the spatial component of multilateral attractiveness is adequately controlled for. However, whether the location fixed effects γ_{ij} will suffice to account for all sources of spatial correlation remains an empirical question (which can be tested; see below).[12] Conversely, if the errors are not i.i.d, the model in Eq. (6) will fail to identify causal relationships.

Several points are worth mentioning here. First, fixed effects are typically used to control for unobserved (or unknown) *correlated heterogeneity* due to omitted, *time-invariant* factors. In the present case of a choice set consisting of the municipalities in a relatively small study area, the fixed effects are likely to have an *inherently spatial dimension* (Debarsy and Ertur, 2010)—even if spatial linkages are not expressed in the form of a parameterized function. As immigrants' location choices are unlikely to be random over space, a tendency towards spatial clustering (and heteroskedasticity) can reasonably be expected, since nearby locations are more likely to be substitutes for one another than more distant locations, as, say, the former are more closely linked by commuting patterns.

[11] The group-location fixed effects, γ_{ij}, may also possibly correct for the potential endogeneity of the migrant-stock variables—i.e., when current inflows and pre-existing stocks are *jointly* determined by location-specific factors. This could be particularly relevant in the case of newly arriving Chinese immigrants, where the stock of Chinese immigrants already in place is of a relatively recent date.

[12] Spatial error correlation (and heteroskedasticity) may also be due to slope-parameter heterogeneity (Peeters, 2012). However, investigation of this issue is beyond the scope of the present paper.

Second, by using an identification strategy relying on location fixed effects, we can dispense with the need to specify a spatial weights matrix. We see this as an important advantage over, say, the use of a spatial-error model (see, e.g., Jayet et al., 2010), since many aspects of the metropolitan system are interrelated in such a complex way that modeling the entire spatial dependence structure is a near impossible task (Pinkse and Slade, 2010; see also McMillen, 2010). Moreover, theory has little to say about how to find the "most relevant" spatial weights, hence leaving ample room for arbitrariness. Insofar as the unobserved factors are only slowly changing over time (if not exactly time-invariant), their influence on new immigrants' location choices is mostly absorbed by the fixed effects. Therefore, their time-invariant nature should not be a source of concern, particularly not with a sufficiently short time span (the empirical analysis below considers a 5-year interval).[13]

Third, by introducing a large number of fixed effects, we may run the risk of potentially saturating the model and losing a large amount of identification power (i.e., the model may become too taxing on the data by controlling for too much), since the identification of the model parameters hinges entirely on the "within"-variation of the explanatory variables, rather than their cross-sectional variation. However, as long as there is sufficient time-variation, this should not really pose a problem, and one can still obtain estimates for all the parameters of interest.[14]

Last, given that identification comes from the *changes* over time, the point at issue here is whether the local characteristics across the spatial units display a tendency *to move together* in response to, say, common macroeconomic shocks. When dealing with local determinants across spatial units within a relatively small geographical area, one should see significant *commonalities* of their time-variation. Indeed, it was found that local factors mostly *do* move together (see the results of a sign test (Snedecor and Cochran, 1989) in Appendix C, Table C.2).

In view of the observed co-movements of local attributes, severe omitted-variable biases are likely to occur (endogeneity in the time dimension). Given these commonalities, it is possible to predict the *direction* of such biases. Given that $\hat{\beta} \xrightarrow{d} \beta + \rho_{yv}(\sigma_v/\sigma_y)$, and assuming $\rho_{yv} < 0$ (note, in particular, that the inclusive value/utility enters the multilateral-attractiveness term with a *minus* sign), the direction of the bias depends on the expected *sign* of the local impact. If the sign is positive (attracting more immigrants), there will be a *downward* bias. To see how this works, consider the following simple example. If GDP per capita at location j (considered here as an attractor) is positively correlated (or co-

[13] Of course, using fixed effects has some limitations. By introducing location-specific effects, we relegate to the (composite) error structure prominent features of immigrants' settlement patterns that, ultimately, need to be understood. Worse, as the "within" estimator sweeps out all of the cross-sectional variation, the model cannot identify what is responsible for the greater part of the variation in the number of migrant arrivals across locations. However, if the goal is to identify the partial (causal) effects of a set of local determinants, the use of fixed effects to address concerns over omitted variables (to control for unobserved, spatially correlated effects) is a satisfactory alternative.

[14] The results reported in the empirical analysis below do not seem to justify this concern.

moves) with GDP per capita in other locations k that prospective immigrants perceive as close substitutes for j, the coefficient on GDP per capita will be *downwardly* biased. This kind of bias is likely to arise because an increase in GDP in location j is associated with an improvement in the attractiveness of alternative locations k. If this multilateral-attractiveness effect is not controlled for, the estimated coefficient on GDP at location j picks up the reduction of the number of migrant arrivals in j due to the increased attractiveness of the alternative locations k (i.e., this reduction is erroneously ascribed to GDP per capita in j). Therefore, the estimated local effect of GDP per capita can hardly qualify as a partial (causal) effect. Obviously, the same logic applies to the unobserved (positively correlated) localized shocks affecting immigrants' perceived utilities. Likewise, if the expected sign is negative (repelling more immigrants), there will be an *upward* bias.

To sum up, failing to control for multilateral attractiveness and its spatial component may lead researchers to *understate* (in absolute value) the impact of the observed local attributes on immigrants' location choices, where even *sign reversals* may occur.

4 Empirical analysis

In this section, we set out the empirical model specification. We use a parsimonious, two-period model, which includes a core set of explanatory variables along with a rich structure of fixed effects (or dummy variables). Data sources and the definition of the variables are given in Appendix A. The choice set of possible destinations is limited to the 41 municipalities of the MMA (see base map in Appendix C, Fig. C.1).[15]

Focusing on a spatially narrowly defined choice set means that all destinations are fairly similar, sharing broadly the same geographical position (located at commuting distance from each other) and a number of structural traits. As already mentioned earlier, we consider five broadly defined groups of immigrants (see Appendix C, Table C.3). Conducting the analysis at a more disaggregated level (say, for individual origin countries) would have produced a large number of zero observations.

4.1 Empirical model specification

Our aim is to identify the local determinants of international migrant arrivals in the municipalities of the MMA using the following baseline model specification:[16]

[15] Given the large geographical size of the city of Madrid (relative to the other municipalities in the metropolitan area), there is scope for defining smaller neighborhoods. However, using the number of migrant arrivals (and migrant stocks) for smaller spatial units (districts) was not possible due to data limitations.
[16] It should be noted that the baseline specification of the empirical model presented here is partly the outcome of several preliminary estimations (not reported here) of many other variants of the model. The baseline specification presented here was the one returning the most meaningful results.

$$n_{ijt} = \exp\Big[\beta_1 \ln PD_{jt-1} + \beta_2 \ln GDP_{jt-1}^{pc} + \beta_3 UR_{jt-1} + (\delta_0 + \delta_1 \Delta EMP_{jt-1} + \delta_2 \Delta EMP_{jt-1}^2) \ln EMP_{jt-1}^{pc} + \beta_4 \ln INC_{jt-1}^{pc} + \beta_5 (\ln INC_{jt-1}^{pc})^2 + \beta_6 \ln PTL_{jt-1} + \beta_7 (\ln PTL_{jt-1})^2 + \beta_8 \ln CI_{jt-1} + \textstyle\sum_{i=1}^{5} \mu_i \, MS_{ijt-1} + \alpha_{it} + \theta_{iMt} + \gamma_{ij}\Big] u_{ijt} \quad (7)$$

The unit of observation is the group-destination-year (indexed i, j, t). Immigration data are for two non-adjacent years, 2005 and 2009. Using the end-points of a 5-year interval warrants sufficient time-variation in the economic variables (note that the economic crisis started in 2007). All explanatory variables have been lagged one year (i.e., measured in the year prior to immigration) to mitigate potential simultaneity biases, and because immigrations do not adjust instantaneously to changes in local characteristics.

Basic descriptive statistics for all variables, by group-location (i, j) combination, are provided in Table 2. The dependent variable represents the number of immigrants from group i arriving in destination j in the metro area. Explanatory variables which represent numbers and those expressed in monetary units enter the model in natural-log form. Monetary values have been expressed in real terms (in constant 2008 prices, CPI deflated). Other explanatory variables enter the empirical model as percentages (not proportions). To keep the model manageable, it is assumed throughout the rest of the paper that the influences of the (group-invariant) local factors are constant across immigrant groups.

<Insert **Table 2** about here>

<u>*4.1.1. Demographic and economic factors*</u>

Population density (PD_j) is used as a proxy for local public goods and other urban amenities, including social protection systems, superior health-care services, better schools, and so on.

GDP per capita (GDP_j^{pc}) may indicate prospects of higher wages. Three different variables are included to capture local labor-market conditions. The 3-year average annual employment-growth rate (ΔEMP_j) acts as a proxy for increasing job opportunities. This variable might be suspect, however. That is, even if employment growth is high, this growth might not be important in terms of job *numbers*. To mitigate the consequences of this uncertainty, we include the employment-growth rate *in interaction with* employment (or number of jobs) per capita (EMP_j^{pc}). Furthermore, the unemployment rate (UR_j) is included to capture other, not yet accounted for job-related factors, considering that higher employment-growth rates and/or an increasing number of jobs per capita do not necessarily imply lower unemployment rates.

Finally, potentials for improved living conditions are captured by gross disposable income per capita (INC_j^{pc}). However, higher incomes could also, at least partly, be indicative of higher housing rents and prices.

4.1.2. Accessibility measures

The number of public transportation lines (PTL_j) is intended to indicate a location's accessibility or connectivity with the city of Madrid (and/or other locations in the MMA). Because the number of public transportation lines in a location is larger the closer it is to the central city of Madrid, this variable could represent a measure of *inverse-distance* to the core of the metro area (the simple correlation between the log of the number of public transportation lines and the log of distance to the city of Madrid is -0.665). To the extent that new immigrants prefer to settle in the suburban reaches of the metropolitan area, the perceived utility of a location is expected to decrease monotonically with a larger number of public transportation lines (or shorter distance to the city of Madrid).

The centrality index (CI_j) is defined as $CI_j = \sum_{k=1}^{J} P_k/D_{kj}$, where P_k is the population size in location k, D_{kj} is the geographical distance (in kilometers) between locations k and j, with *intra*-location distance defined as $D_{jj} = 0.67\sqrt{A_j/\pi}$ (in order to avoid "donut holes", the reference location j is also included in the summation). The centrality index measures the attractiveness of a location's relative position within the MMA. That is, a positive sign of the coefficient on the centrality index signifies a tendency of (inward) agglomeration of immigrants towards the core of the metropolitan area, whereas a negative sign indicates a tendency of (outward) de-agglomeration.

4.1.3. Migrant stocks

We also include the local "migrant stocks", measured as the percentage of the resident immigrant community for each group i of the total population in location j; that is, $MS_{ij} = V_{ij}/P_j \times 100$, where V_{ij} is the stock (absolute number) of previous immigrants from group i residing in j (which may also include immigrants who re-migrated within Spain), and P_j is the total population in j.

The size of the migrant stock is usually expected to have a positive effect on a location's perceived utility, as established networks of "family and friends" can provide prospective immigrants with information about economic conditions, support in managing the immigration process, and help in obtaining housing and finding a job. However, as already mentioned before, the migrant-stock effect may turn out to be insignificant because of adverse selection, hetero-localism, and the heterogeneity of broadly defined groups of immigrants.

4.1.4. Fixed effects

A rich structure of fixed effects enters the empirical model, including group-location fixed effects, group-year dummies, and group-Madrid-year dummies.

The group-location fixed effects (γ_{ij}) are intended to control for the unobservable (relatively stable) spatial-structural components of multilateral attractiveness, which are allowed to be different across immigrant groups (heterogeneous preferences). So, we seem to go one step further than Guimarães et al. (2004), who introduced fixed effects for each location only ($\gamma_{ij} = \gamma_j$). The group-year dummies (α_{it}) are included to ensure compatibility of the Poisson regression with the RUM-consistent CL model. Finally, group-Madrid-year dummies (θ_{iMt}) are added to capture the "idiosyncratic nature" of the city of Madrid—even if only because of the scale effect (the city of Madrid attracts around 50% of all new immigrants to the MMA).

4.2 Parameter estimates

In this section, we present the results from the estimation of the fixed-effects Poisson model. We begin with a discussion of the parameter estimates obtained from the baseline model in Eq. (7). A summary of the estimation results is presented in Table 3. This discussion is followed by an analysis of the residuals returned by the preferred Poisson model. Finally, we expand on the intrinsic uncertainty surrounding the (semi-)elasticities derived from Poisson.

<Insert **Table 3** about here>

4.2.1. Comparison of estimates across model specifications

A comparison of the results across the columns of Table 3 immediately reveals that the parameters estimates are strongly sensitive to the choice of model specification. A first point to notice is that the coefficients remain virtually unchanged by the introduction of the group-year dummies, α_{it}. In many instances, the estimates in columns 1–2 have an unexpected sign, hence giving rise to misleading inferences. It is only after incorporating the (time-invariant) location-specific effects, γ_{ij}, that estimates are markedly different and change in the expected direction, not only in *magnitude* (in accordance with our predictions from the influences of multilateral attractiveness), but also in *sign*.

These findings highlight the "contextual (or situational)" nature of multilateral attractiveness. They also illustrate two outstanding features of the fixed-effects estimation in the present application. First, controlling for correlated spatial effects is crucially important for obtaining meaningful results (note also, in passing, that the Hausman test strongly rejects the null hypothesis of "no correlated effects").[17] Second, the introduction of location fixed effects does not weaken the identification power of the

[17] Note that the fixed-effects and the random-effects estimators share the common feature of returning (group-wise) individual "Madrid effects" that match with the results of the exploratory shift-share analysis (i.e., "area"-shift component; see Appendix C, panel A of Table C.1).

within-variation of the local attributes (by absorbing too much of the useful variation of the data). In fact, we find quite the opposite.

Intuitively, the large impact that the conditioning on location-specific effects has on the estimates suggests that spatial patterns of immigration are quite *persistent* over the time span considered here; that is, locations initially receiving large (small) numbers of immigrants are likely to maintain high (low) numbers five years later. To corroborate our intuition, we computed the within and between variation of observed immigration numbers, and found that the between variation is about 10 times larger than the within variation ($CV_{between} = 4.6$ and $CV_{within} = 0.437$). This means that, not surprisingly, immigration numbers vary much more *across* locations, as opposed to *within* locations. Next, we investigated the role that initial (2005) immigration patterns play in determining subsequent (2009) immigration patterns—i.e., the extent to which *future* location choices are related to location choices in the *past*. To measure the impact of the initial choices, we included the (log) immigration numbers in 2005 as an additional explanatory variable and estimated a cross-section model for 2009. We found (see results in Appendix C, Table C.4) that the estimated coefficient on 2005 immigrations is highly significant (t = 14.83) and among the largest determinants of immigrations in 2009, returning an "elasticity" of 0.75. Even though this elasticity is statistically smaller than one (the 95% confidence interval was estimated at 0.65–0.85), this outcome emphasizes that "historical" immigration patterns play an vital role in determining future patterns, even after controlling for local sources of variation (including the size of local migrant stocks).

In view of these general findings, the discussion that follows will be focused only on the estimates reported in column 4 of Table 3 (results for the baseline model specification).

4.2.2. Parameter estimates for the baseline specification

We now provide a somewhat more detailed discussion of the estimates obtained for the baseline specification reported in column 4 of Table 3. The estimated partial effects for the variables that enter the model in a non-linear form (i.e., quadratic and/or interaction terms) are depicted in Fig. 2, to ease their interpretation.

The coefficient on population density is positive, suggesting immigrants' preferences for choosing urbanized places, likely because of local amenity-related motives, including easier access to public (social) services. GDP per capita appears to have no significant effect on a location's attractiveness. A possible explanation for this finding is that immigrants may have a preference for settling in a location (place of residence) at commuting distance of their workplace, where the major activities take place and higher wages are paid, possibly because immigrants may find living next to their workplace unattractive, or too costly.

<Insert **Fig. 2** about here>

The estimation results for the local labor-market conditions are somewhat mixed, yet they appear to be in line with earlier studies that come to different conclusions as to whether labor-market conditions affect immigrants' location choices (Zavodny, 1999; Åslund, 2005). Locations with higher unemployment rates are found to lose immigrants, although the negative effect is only marginally significant. Dodson (2001, p. 53) provides a credible explanation for this weak effect to occur, namely the unemployment rate in relation to the source country's unemployment rate is more important, in the sense that every location in the MMA may have an unemployment rate that is so low compared to the source country's unemployment rate (in most instances) that this variable makes little difference in the location decision. On the other hand, the effect of employment per capita is found to be positive and significant, and generally tends to be generally stronger if the initial employment-growth rate (in the 3-year period prior to the arrival of new immigrants) is higher (see Panel A of Fig. 2). Somewhat surprisingly, the effect weakens at an initial job-growth rate of 5.6% or higher. A possible explanation could be that other factors could be at play. For example, strongly improved local job opportunities may not be fully anticipated by prospective immigrants or capitalized into higher wages (Partridge et al., 2008). Also, improved labor-market expectations may not necessarily imply more job opportunities for new immigrants if there are adverse *industry-mix* effects (Partridge and Rickmann, 1999) or *job-skills mismatches* (OECD, 2007). To put it somewhat more bluntly: the most favorable labor-demand shocks may have occurred in the "wrong" economic activities, so marginal (perceived) disutility may arise from this mismatch. It is evident that more research is needed to find out whether and to what extent our results are related to the industrial structure of job-growth across locations. However, several data limitations prevent us from carrying forward work on this issue.

The coefficients on gross disposable income per capita and its square, positive and negative, respectively, suggest that locations become less attractive with increasing income levels—eventually turning into a negative (repulsive) effect at very high income levels. A possible explanation is that high income levels may indicate a local shortage of dwellings at affordable prices. Thus, housing costs may offset, or outweigh, the benefits of improved living conditions (see panel B of Fig. 2). Due to the lack of relevant municipality-level data, it was not possible to include housing costs as an additional explanatory variable.

The results related to the accessibility measures all point in the same direction, namely that of suburbanization. First, the coefficients on the number of public transportation lines and its square, positive and negative, respectively, suggest that the perceived utility of a locality decreases with greater accessibility or connectivity with the city of Madrid and other locations in the metro area (see panel C of Fig. 2). This result amply reflects the tendency of new immigrants to settle in dispersed areas, providing substance to the hypothesis that transportation encourages urban sprawl towards to suburban fringe of the MMA (see also Garcia-López, 2012).[18] Second, the coefficient on the centrality index is negative and significant, indicating an increasingly dispersed settlement pattern towards non-central suburban areas. Many concurrent de-glomeration forces could be at work, making core areas less attractive, such as high costs (rent, services, taxes), congestion (traffic), environmental externalities (pollution, noise), and perhaps also a higher incidence of crime—besides the continued development of real-estate projects in Madrid's satellite areas (with apartments becoming increasingly available at affordable prices and/or favorable mortgage terms).

Finally, we turn to the impact of the size of local migrant stocks. The attractiveness of the presence of immigrant communities in a location varies considerably from one group to another. The migrant-stock coefficient is *positive* and significant only for Chinese immigrants. A possible explanation for this finding is that Chinese immigrants understand the importance of strong local (*guanxi*) networks of "family and friends" and mutual trust relationships (the so-called *Wenzhou* model; e.g., Mallet, 2011), which, together with the wide variety of economic activities they are involved in (retail/wholesale, import/export, real estate, services, etc.), may explain their remarkable resilience to the economic crisis (e.g., Bilefsky, 2013). In the case of Bulgarian–Romanian immigrants, the coefficient is *negative* and significant, suggesting possible in-group rivalry (see also, e.g., Serban and Voicu, 2010). Many of them are low-skilled people traditionally seeking—mostly temporal—jobs in sectors (e.g., metallurgical industry, construction, etc.) that are hard-hit by the economic downturn. Owing to the increasingly unfavorable labor-market conditions, they are likely to encounter fierce competition for jobs from the local immigrant community already in place (making it increasingly difficult to find a job at their reservation wage), which may therefore encourage them to look for non-traditional destinations (i.e., away from saturated network destinations) in search for better job opportunities.[19] On the other hand, the coefficient is *insignificant* for both Latin-American and EU25 immigrants. One (obvious) reason for this insignificant result is the concealed heterogeneity of these broadly defined groups. In addition, Latin-Americans immigrants may not feel an urgency of living in close-knit *latino* communities, as their innate Spanish-language proficiency makes it evidently easier for them to assimilate in the wider environment. For their part, EU25 immigrants are likely to have a broader range of residential possibilities. It is thinkable that, given their relatively privileged position (e.g., in terms of income and educational level), many of them can afford to settle anywhere in the metropolitan area (say, in the most attractive suburbs). Last, the case of Moroccan immigrants, where we find only a *marginally significant* coefficient, seems to be somewhat more complicated, prompting questions for which we have no ready answers at this point. Because of the specificity of their ethnic amenities (e.g., easy

[18] This result contrasts sharply with the findings in Bayona-Carrasco and Gil-Alonso (2012) with regard to the settlement patterns of immigrants to the Barcelona metropolitan area.
[19] Such responses would run counter to those predicted by Patel and Vella (2013).

access to a mosque, *halal* food, etc.), one would have logically expected to find a tendency towards clustering in ethnic neighborhoods (e.g., Gest, 2010).[20] Yet, as mentioned before, a dispersal of their settlement (hetero-localism) should not prevent the newly arriving Moroccans from sustaining strong community ties through various communication means and socio-cultural activities at certain places.

4.3 Analysis of residuals (diagnostics)

Controlling for multilateral attractiveness is quite challenging, since it requires the residuals of the Poisson model to be cross-sectionally independent. This calls therefore for a careful investigation of those residuals.

Different types of residuals can be computed for non-linear model such as Poisson. A natural starting point is to look at the *raw* residuals, $\hat{r}_{ijt} = n_{ijt} - \hat{n}_{ijt}$, where $n_{ijt} = e^{\hat{\gamma}_{ij}+\hat{\alpha}_{it}+\hat{\theta}_{iMt}+\hat{\boldsymbol{\beta}}_1'\mathbf{y}_{jt}+\hat{\boldsymbol{\beta}}_2'\mathbf{x}_{ijt}}$. We also check the pattern of the *Pearson* residuals, which "correct for" the implied heteroskedasticity. The Pearson residuals, $\tilde{r}_{ijt}$, can be obtained by dividing the raw residuals by their standard error, such that $\tilde{r}_{ijt} = \hat{r}_{ijt}/SE(\hat{n}_{ijt})$. From the Poisson variance assumption, it follows that $\tilde{r}_{ijt} = \hat{r}_{ijt}/\sqrt{\hat{n}_{ijt}}$ (Wooldridge, 2002, p. 462). So, the scaling of the raw residuals puts them on an "equal footing" in terms of variance. Therefore, if the Poisson variance assumption holds, the variability of the Pearson residuals should be fairly constant (homoskedastic) when plotted against the fitted values.

4.3.1. Recovering the group-destination fixed effects

Before proceeding, we first need to compute the residuals from the estimated Poisson regression. It has been shown by Baltagi (2009, p. 230) that the group-location fixed effects, $\hat{\gamma}_{ij}$, can be recovered from the total counts as

$$\hat{\gamma}_{ij} = \ln\left[\frac{\sum_{t=1}^{T} n_{ijt}}{\sum_{t=1}^{T} e^{\hat{\alpha}_{it}+\hat{\theta}_{iMt}+\hat{\boldsymbol{\beta}}_1'\mathbf{y}_{jt}+\hat{\boldsymbol{\beta}}_2'\mathbf{x}_{ijt}}}\right] \tag{8}$$

from which the raw residuals, $\hat{r}_{ijt}$, can easily be derived. In what follows, we focus on the 2009 raw residuals, since having only two time periods, it follows that $\hat{r}_{ij,2009} = -\hat{r}_{ij,2005}$ (see proof in Appendix B). Note, however, that this relationship does not generally hold for the Pearson residuals, $\tilde{r}_{ijt}$.

4.3.2. Testing the Poisson variance assumption

A potential problem with the fixed-effects Poisson regression is that the count data frequently suffer from over-dispersion (e.g., Allison and Waterman, 2002). To assess the adequacy of our Poisson regression, we test its variance assumption using the modified Park test proposed by Manning and

---

[20] Moroccan immigrants do appear to settle, though, in ethnic communities *within* the city of Madrid (and in some surrounding municipalities). Unfortunately, the present empirical analysis does not allow us to uncover this kind of settlement pattern.

Mullahy (2001, p. 471). Considering a "power-proportional" form of the mean-variance structure, $V(n_{ijt}|\bullet) = \lambda_0[E(n_{ijt}|\bullet)]^{\lambda_1} = \lambda_0(e^{\gamma_{ij}+\alpha_{it}+\theta_{iMt}+\beta_1'\mathbf{y}_{jt}+\beta_2'\mathbf{x}_{ijt}})^{\lambda_1}$, the Park test allows us to check whether the conditional variance is proportional to the conditional mean.

We begin with a standard approach using an auxiliary gamma GLM regression with log link (with a non-robust variance estimator), based on the raw residuals (note that we omit the subscript t, as our focus is on the 2009 residuals only):

$$\hat{r}_{ij}^2 = e^{\ln\lambda_0+\lambda_1\ln\hat{n}_{ij}}\varepsilon_{ij}, \tag{9}$$

and test the null hypotheses that $\widehat{\lambda}_1 = 1$ (variance proportional to mean) and $\widehat{\lambda}_0 = 1$ (equidispersion/ no overdispersion).

Next, we use the test proposed by Santos Silva and Tenreyo (2006, p. 646) based on a first-order Taylor-series expansion of $\lambda_0\hat{n}_{ijt}^{\lambda_1}$ around $\lambda_1 = 1$, and applied to the Pearson residuals, $\tilde{r}_{ijt} = \hat{r}_{ijt}/\sqrt{\hat{n}_{ijt}}$, using the following auxiliary regression:

$$\tilde{r}_{ijt}^2 = \lambda_0 + \lambda_0(\lambda_1 - 1)\ln\hat{n}_{ijt} + \xi_{ijt} \tag{10}$$

and test the null hypotheses that $\widehat{\lambda}_0(\widehat{\lambda}_1 - 1) = 0$ or $\widehat{\lambda}_1 = 1$ (variance proportional to mean/homoskedasticity of the Pearson residuals) and $\widehat{\lambda}_0 = 1$ (equidispersion/no overdispersion).

The results returned by these two tests, presented in Table 4, show that the Poisson mean-variance assumption cannot be rejected, at least not for the immigrant groups taken separately.

<Insert **Table 4** about here>

4.3.3. Testing for spatial error autocorrelation

To test for the presence of cross-sectional (spatial) error dependence, we perform the standard Moran's I test, using the raw and Pearson residuals from the estimated Poisson model.[21] Specifically, we test whether the group-destination fixed effects have been able to account for the unobserved spatial structure of multilateral attractiveness (i.e., through complex interactions linking all locations in the metro area). We test for the *within-group* spatial autocorrelation among the residuals for each immigrant group separately.

[21] Moran's I statistic is found to be very powerful in detecting spatial error correlation (besides misspecification of the model), even in small samples (Kelejian and Robinson, 1998, p. 391; Anselin, 2005, p. 197). Therefore, when not statistically significant, the Moran's I test accepts the null of "no spatial autocorrelation". Yet, the outcomes of the Moran's I test may be sensitive to the specification of the spatial weight matrix (Hsiao et al., 2012). Since there is little consensus about the most appropriate choice of the spatial weights matrix, we use a simple first-order contiguity matrix. For testing residual spatial autocorrelation, a complete and correct (theory-based) specification of the spatial relationships is not generally necessary. Although a correct specification yields a powerful, consistent test, even tests against misspecified alternatives generally pick up some of the spatial dependence, albeit with a possibly significant loss of power (Pinkse and Slade, 2010, p. 112).

The overall picture provided by the test results, summarized in Table 5, suggests that both the raw and Pearson residuals returned by the fixed-effects Poisson regression display no discernible pattern of (global) spatial autocorrelation, so the residuals appear to be in accordance with the IIA assumption underlying the CL model. This means that the introduction of location fixed effects has been effective in establishing cross-sectional (spatial) independence of the remaining unobserved utility components (see results in panel B of Table 5). In contrast, the inclusion of the group(-year)-specific effects alone is clearly insufficient to restore cross-sectional error independence if multilateral attractiveness has a spatial dimension (see the results in panel A of Table 5).

<Insert **Table 5** about here>

4.4 Uncertainty surrounding the implied (semi-)elasticities

A final issue relates to the built-in uncertainty surrounding the reported Poisson estimates. Despite the equivalence relation between CL and Poisson, the implied elasticities, or semi-elasticities, have a different quantitative interpretation. Specifically, CL and Poisson provide a *conservative* (lower-bound) and a *liberal* (upper-bound) quantity, respectively, as was shown by Schmidheiny and Brülhart (2011). That is, CL concords with *zero-sum* reallocations of immigrants across destinations (providing *conditional* immigration responses, given a *fixed* total number of immigrants), whereas Poisson accords with *positive-sum* reallocations of immigrants across destinations (providing *unconditional* immigration responses, with a *varying* total number of immigrants).[22]

Using broadly similar notation as in Schmidheiny and Brülhart (2011, p. 218), the CL interpretation of the *proportionate* change in the expected number of new immigrants following a 1-unit change in a given local attractor (*semi-elasticity*) is:

$$\frac{\partial \ln E(n_{ij})}{\partial x_j} = \beta_j(1 - P_{ij}) \tag{11}$$

where β_j is the Poisson value, and $P_{ij} = n_{ij}/n_i$. Thus, the larger (smaller) the initial proportion, P_{ij}, the smaller (larger) the CL lower bound—or, the larger (smaller) the discrepancy between the upper and

[22] Intermediate cases between the two extremes can be represented by a Nested Logit (NL) model featuring an "outside option" (Schmidheiny and Brülhart, 2011). For ease of interpretation, one may think of the outside option as covering "the rest of Spain", representing all locations outside the MMA, jointly denoted by $j = 0$ (the locations within the MMA are jointly denoted by $j > 0$). Then,

$$\partial \ln E(n_{ij})/\partial x_j \,|_{j>0} = \beta_j\left[1 - P_{ij|j>0}(1 - \lambda P_{i0})\right] = \beta_j(1 - \rho P_{ij|j>0})$$

where λ is the "rivalry parameter" ($0 < \lambda \leq 1$), and $\rho = 1 - \lambda P_{i0}$ can be thought of as capturing the relative importance of the MMA vis-à-vis the rest of Spain. Thus, the point at issue here is how to assess the potential demand substitution that can take place when immigrants on the margin between the MMA ($j > 0$) and the rest of Spain ($j = 0$) affect the distribution of immigrants in Spain (at the country level). Investigation of this issue is beyond the scope of the present paper, though it can reasonably be expected that the "true" quantities are somewhere in the middle between the two extremes.

lower bounds.[23] Along similar lines, and following Brülhart et al. (2012, p. 1089), the expected change in the *choice probability* is:

$$\frac{\partial P_{ij}}{\partial x_j} = \beta_j P_{ij}(1 - P_{ij}) \tag{12}$$

With a relatively small number of observational units (J = 41), the quantitative differences can be significant for large destinations. Hence, the uncertainty about the "true" value of the *semi-elasticity* for the city of Madrid (with an average initial proportion, for all immigrant groups taken together, of 0.5) can be quite substantial. However, if we look at the *average* effects, the differences are generally small, or even negligible: the average proportion, for all groups taken together, is only 0.024 (or 0.013 if the city of Madrid is excluded). This means that the CL lower bounds of the (semi-)elasticities are *on average* only 2.4% (1.3%) below the Poisson upper bounds.

To get a sense of the quantitative differences between CL and Poisson, we take the local migrant-stock effect, ceteris paribus, for Chinese immigrants as an illustrative example. The most important difference occurs, as expected, for the city of Madrid (with 3.2 million inhabitants and a 2008 stock of 24,000 Chinese in 2009), which receives the lion's share of the total number of Chinese immigrants to the MMA (56.9% in 2009). The CL interpretation of the Poisson value (estimated at 1.283, see column 4 of Table 3) implies that, say, a 0.1%-point increase in the stock of Chinese immigrants (relative to the total population) increases the probability of choosing the city of Madrid as the destination by 2.4% points (= 0.1 × (1.283 − 0.294) × 0.569 × (1 − 0.569) × 100), after correction for the individual "Madrid effect" (estimated at −0.294), whereas Poisson predicts an increase in the choice probability of 6.5% points (= 0.1 × (1.283 − 0.294) × 0.569 × 100, if evaluated at the total number of new Chinese immigrants in 2009.

Table 6 shows the CL and Poisson values of the expected change in the choice probabilities and the number of new arrivals of Chinese immigrants for a selection of municipalities in the MMA (see the base map in Appendix C, Fig. C.1), following a 1,000 addition to the local migrant stock and a 0.1%-point increase in the local stock, respectively. An inspection of the results reveals that the quantitative differences are only of minor importance for Parla (with 108,000 inhabitants and a stock of about 1,000 Chinese immigrants in 2009) or Fuenlabrada (with 195,000 inhabitants and a stock of about 1,100 Chinese immigrants in 2009), and negligible for Tres Cantos (with 41,000 inhabitants and a stock of only about 100 Chinese immigrants in 2009).

Based on these findings, it is fair to conclude that the more conservative CL quantities do not fundamentally alter the conclusions of our empirical analysis in a *qualitative* sense.

---

[23] Note further that the *cross*-(semi-)elasticity in the positive-sum, Poisson case, $\partial \ln E(n_{ik})/\partial x_j$ is equal to zero, as shown by Schmidheiny and Brülhart (2011, Table 2, p. 218). This means that locations are non-rivaling, in the sense that one location's gain in terms of immigrants is not another location's loss.

< Insert **Table 6** about here >

5 Summary and concluding remarks

The main interest of this paper was centered on the potential problems caused by the omission of the influence exerted by the (unobserved) multilateral attractiveness on immigrants' location choices. The primary intent of the empirical analysis was to illustrate how the introduction of both group-specific and location-specific fixed effects in a Poisson model can be useful to cope with the issues of both multilateral attractiveness and spatial structure, thereby strengthening the identification power of the estimation while maintaining consistency with the theoretical RUM framework.

We showed that multilateral attractiveness matters in the case of international migration to the MMA, where the immigrant's choice of a particular location depends on "what happens in the rest of the metro area". As unilateral predictions do not readily extend to a multilateral world with complex interactions linking all locations in the metropolitan area, the inclusion of a multilateral-attractiveness term is crucial for the econometric results one obtains. Moreover, it was shown that multilateral attractiveness is a contextual (or situational) phenomenon rather than a global one; that is, geography seems to dictate the substitutability of locations. That is, the (unobserved) contextual features play a more prominent role for locations that are in close proximity to one another. To deal with the issue of spatial-structural effects, our model included location-specific fixed effects, and we found that cross-sectional error independence is ensured. The empirical analysis manifestly illustrated that neglecting the multilateral attractiveness of all locations, along with its inherently spatial dimension, inevitably leads to conclusions that are completely at odds with those predicted by models that *do* control for it.

Regarding specifics, it was found that, ceteris paribus, local economic conditions such as GDP per capita and the unemployment rate do not really matter in choosing a particular location whereas most other local characteristics do. On the other hand, the utility obtained from the presence of an ethnic community in a location turns out to be markedly different across immigrant groups. The migrant-stock coefficient is positive and strongly significant only for Chinese immigrants (suggesting positive network externalities), negative and significant for Bulgarian–Romanian immigrants (suggesting in-group rivalry), and insignificant for EU25 and Latin-American immigrants (suggesting hetero-local settlement preferences). For Moroccan immigrants, however, the picture is somewhat less clear.

Although the results presented in this paper are informative, the analysis can be improved along several dimensions. For instance, we assumed homogeneous compositions of the immigrant groups, e.g., in terms of skill levels. It would be interesting to investigate the incidence of (positive/negative) selection and sorting. Unfortunately, at this point no relevant data were available at the level of the municipalities. Yet we consider this an interesting avenue for future research.

Acknowledgements

The research for this paper was partly conducted during a sabbatical leave by the first author at the Lawrence R. Klein Institute, Autonomous University of Madrid, which was funded by the Research Foundation–Flanders (Grant No. K801112N). The second author acknowledges financial support from the Spanish Ministry of Education and Science (Grant No. ECO2009-10534) and the Xunta de Galicia (Grant No. 10SEC201032PR).

References

Allison, P. D., Waterman, R.P., 2002. Fixed-effects negative binomial regression models, Sociological Methodology 32, 247–265.

Anderson, J.E., van Wincoop, E., 2003. Gravity with gravitas: A solution to the border puzzle, American Economic Review 93, 170–192.

Anselin, L., 2005. Exploring Spatial Data with GeoDa™: A Workbook, Spatial Analysis Laboratory and Center for Spatially Integrated Social Science, Dept. of Agriculture and Consumer Economics, University of Illinois, Urbana-Champaign, IL.

Arauzo-Carod, J.M., Liviano-Solis, D., Manjón-Antolín, M., 2010. Empirical studies in industrial location: An assessment of their methods and results, Journal of Regional Science 50, 685–711.

Åslund, O., 2005. Now and forever? Initial and subsequent location choices of immigrants, Regional Science and Urban Economics 35, 141–165.

Baltagi, B.H., 2009. A Companion to Econometric Analysis of Panel Data, John Wiley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex, England.

Bauer, T., Epstein, G.S., Gang, I.N., 2007. The influence of stocks and flows on migrants' location choices, Research in Labor Economics 26, 199–229.

Bayona-Carrasco, J., Gil-Alonso, F., 2012. Suburbanisation and international immigration: The case of the Barcelona Metropolitan Region (1998–2009), Journal of Economic and Social Geography 103, 312–329.

Bertoli, S., Fernández-Huertas Moraga, J., 2012. Visa policy, networks and the cliff at the border, Discussion Paper Series IZA DP No. 7094, December.

Bertoli, S., Fernández-Huertas Moraga, J., 2013. Multilateral Resistance to Migration, Journal of Development Economics, forthcoming, doi: 10.1016/j.jdeveco.2012.12.001.

Bilefsky, D., 2013. Spain's Chinese Immigrants Thrive in Tough Economy, The New York Times, January 2, 2013, available at http://www.nytimes.com/2013/01/03/world/europe/spains-chinese-immigrants-thrive-in-tough-economy.html?_r=0.

Borgers, A., Timmermans, H.J.P., 1988. A context-sensitive model of spatial choice behaviour, in: Gollodge, R.O., Timmermans, H.J.P. (Eds.), Behavioural Modelling in Geography and Planning, Croom Helm, London, pp. 159–178.

Brown, W.M., Scott, D.M., 2012. Human capital location choice: Accounting for amenities and thick labor markets, Journal of Regional Science 52, 733–761.

Brülhart, M., Jametti, M., Schmidheiny, K., 2012. Do agglomeration economies reduce the sensitivity of firm location to tax differentials?, The Economic Journal 122, 1069–1093.

Cadena, B.C., 2013. Native competition and low-skilled immigrant inflows, Journal of Human Resources forthcoming, available at SSRN: http://ssrn.com/abstract=1619001.

Clark, X., Hatton, T.J., Williamson, J.G., 2007. Explaining U.S. immigration, 1971-1998, The Review of Economics and Statistics 89, 359–373.

Davies, P.S., Greenwood, M.J., Li, H., 2001. A Conditional Logit approach to U.S. state-to-state migration, Journal of Regional Science, 41, 337–360.

Debarsy, N., Ertur, C., 2010. Testing for spatial autocorrelation in a fixed effects panel data model, Regional Science and Urban Economics 40, 453–470.

Dodson, M.E., 2001. Welfare generosity and location choices among new United States immigrants, International Review of Law and Economics 21, 47– 67.

Edin, P.A., Frederiksson, P., Åslund, O., 2003. Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment, Quarterly Journal of Economics 118, 329–357.

Gang, I.N., Rivera-Batiz, F.L., 1994. Labor market effects of immigration in the United States and Europe: Substitution vs. complementarity, Journal of Population Economics 7, 157–175.

Garcia-López, M.A., 2012. Urban spatial structure, suburbanization and transportation in Barcelona, Journal of Urban Economics 72, 176–190.

Gest, J. 2010. Apart: Alienated and Engaged Muslims in the West, Hurst & Co., London, UK.

Grzymala-Kazlowska, A., 2005. From ethnic cooperation to in-group competition: Undocumented Polish workers in Brussels, Journal of Ethnic and Migration Studies 31, 675–697.

Guimarães, P., Figueiredo, O., Woodward, D., 2003. A tractable approach to the firm location decision problem, The Review of Economics and Statistics 85, 201–204.

Guimarães, P., Figueiredo, O., Woodward, D., 2004. Industrial location modeling: Extending the Random Utility Model, Journal of Regional Science 44, 1–20.

Hanson, G.H., 2010. International migration and the developing world, in: Rodik, D., Rozenzweig, M. (Eds.), Handbook of Development Economics, vol. 5, North-Holland, The Netherlands, pp. 4363–4414.

Hardwick, S.W., 2006. Nodal heterolocalism and transnationalism at the United States–Canadian border, The Geographical Review 96, 212–228.

Hardwick, S.W., Meacham, J.E., 2005. Heterolocalism, networks of ethnicity, and refugee communities in the Pacific Northwest: The Portland story, The Professional Geographer 57, 539-557.

Hatton, T.J., Williamson, J.G. 2011. Are Third World emigration forces abating?, World Development 39, 20–32.

Hsiao, C., Pesaran, M.H., Pick, A., 2012. Diagnostic tests of cross-section independence for limited dependent variable panel data models, Oxford Bulletin of Economics and Statistics 74, 253–277.

Jaeger, D.A., 2008. Green cards and the location choices of immigrants in the United States, 1971 –2000, Research in Labor Economics 27, 131–183.

Jayet, H., Ukrayinchuk, N., De Arcangelis, G., 2010. The location of immigrants in Italy: Disentangling networks and local effects, Annals of Economics and Statistics/Annales d'Économie et de Statistique 97/98, 329–350.

Kaushal, N., 2005. New immigrants' location choices: Magnets without welfare, Journal of Labor Economics 23, 59–80.

Kelejian, H.H., Robinson, D.P., 1998. A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results, Regional Science and Urban Economics 28, 389–417.

Liu, C.Y., 2013. Latino immigration and the low-skill urban labor market: The case of Atlanta, Social Science Quarterly 94, 131–157.

Mallet, V. 2011. Chinese arrivals take Spain by storm, The Financial Times, January 21, 2011, available at www.ft.com/intl/cms/s/0/cccbd64c-258d-11e0-8258-00144feab49a.html#axzz2JsfVf75y.

Manning, W.G., Mullahy, J., 2001. Estimating log models: To transform or not to transform?, Journal of Health Economics 20, 461–494.

Mayda, A.M., 2010. International migration: A panel data analysis of the determinants of bilateral flows, Journal of Population Economics 23, 1249–1274.

McFadden, D.L., 1974. "Conditional Logit analysis of qualitative choice behavior, in: Zarembka, P. (Ed.), Frontiers in Econometrics, Academic Press, New York, pp. 105–142.

McFadden, D.L., 1984. Econometric analysis of qualitative response models, in: Grilliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. 2, Elsevier Science Publishers, Amsterdam, pp. 1396–1457.

McKenzie, D., Theoharides, C., Yang, D., 2012. Distortions in the international migrant labor market: Evidence from Filipino migration and wage responses to destination country economic shocks, Policy Research Working Paper No. 6041, The World Bank.

McMillen, D.P. 2012. Issues in spatial data analysis, Journal of Regional Science 50, 119 –141.

Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the U.S. labor market, Quarterly Journal of Economics 118, 549–597.

Newbold, K.B., 1999. Spatial distribution and redistribution of immigrants in the metropolitan United States, 1980 and 1990, Economic Geography 75, 254–271.

Newbold, K.B., Spindler, J., 2001. Immigrant settlement patterns in metropolitan Chicago, Urban Studies 38, 1903–1919.

OECD, 2007. OECD territorial reviews: Madrid, Spain. Organisation for Economic Co-operation and Development, Paris.

Ortega, F., Peri, G., 2009. The causes and effects of international migrations: Evidence from OECD countries 1980–2005, NBER Working paper No. 14883, National Bureau of Economic Research.

Partridge, M.D., Rickman, D.S., 1999. A note on the benefits to current residents of state employment growth: Is there an industry mix effect on migration?, Journal of Regional Science 39, 167–181.

Partridge, M.D., Rickman, D.S., Ali, K., Olfert, M.R., 2008. Lost in space: Population growth in the American hinterlands and small cities, Journal of Economic Geography 8, 727–757.

Patel, K., Vella, F., 2013, Immigrant networks and their implications for occupational choice and wages, forthcoming in The Review of Economics and Statistics.

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.

Peeters, L. 2012. Gravity and spatial structure: The case of interstate migration in Mexico, Journal of Regional Science 52, 819–856.

Pellegrini, P.A., Fotheringham, A.S., 2002. Modelling spatial choice: A review and synthesis in a migration context, Progress in Human Geography 26, 487–510.

Pinkse, J., Slade, M.E., 2010. The Future of spatial econometrics, Journal of Regional Science 50, 103–117.

Santos Silva, J.M.C., Tenreyo, S., 2006. The log of gravity, The Review of Economics and Statistics 88, 641–658.

Schmidheiny, K., Brülhart, M., 2011. On the equivalence of location choice models: Conditional Logit, Nested Logit and Poisson, Journal of Urban Economics 69, 214–222.

Scott, D.M., Coomes, P.A., Izyumov, A.I., 2005. The location choice of employment-based immigrants among U.S. metro areas, Journal of Regional Science 45, 113–145.

Serban, M., Voicu, B., 2010. Romanian migrants to Spain: In- or outside the migration networks a matter of time?, Revue d'Études Comparatives Est-Ouest 41, 97–124.

Snedecor, G.W., Cochran, W.G., 1989. Statistical Methods, eight ed. Iowa State University Press, Ames, IA.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge, MA.

Zavodny, M., 1999. Determinants of recent immigrants' locational choices, International Migration Review 33, 1014–1030.

Zelinsky, W., Lee, B.A., 1998. Heterolocalism: An alternative model of the sociospatial behaviour of immigrant ethnic communities, International Journal of Population Geography 4, 281–298.

Table 1
Immigrations to the Madrid metropolitan area and the city of Madrid.

	2005		2009		2009–2005	2005	2009
A: *Metropolitan area*							
	Number	%	Number	%	%Difference	% of Region	% of Region
EU25	12316	7.3%	13262	8.4%	7.7%	91.1%	92.2%
Bulgaria–Romania	36223	21.3%	22700	14.5%	–37.3%	88.3%	86.6%
Latin America	75938	44.7%	72995	46.5%	–3.9%	94.9%	94.0%
Morocco	11711	6.9%	12434	7.9%	6.2%	82.3%	82.1%
China	7399	4.4%	9900	6.3%	33.8%	98.5%	96.5%
Subtotal	*143587*	*84.6%*	*131291*	*83.6%*	*–8.6%*	*91.9%*	*91.4%*
Other origins	26128	15.4%	25766	16.4%	–1.4%	95.7%	95.8%
Total	169715	100.0%	157057	100.0%	–7.5%	92.4%	92.1%
B: *City of Madrid (only five immigrant groups)*							
	Number	% of number in metro area	Number	% of number in metro area			
EU25	6585	53.5%	8104	61.1%			
Bulgaria–Romania	12574	34.7%	6967	30.7%			
Latin America	45416	59.8%	40249	55.1%			
Morocco	4211	36.0%	3643	29.3%			
China	4846	65.5%	5629	56.9%			
Total	73632	51.3%	64592	49.2%			

Table 2
Summary statistics—Variables included in the empirical model.

	2005					2009				
	Mean	S.D.	Min.	Median	Max.	Mean	S.D.	Min.	Median	Max.
A: *Number of new immigrants*, by group-location combination										
EU25	300.4	1015.8	8	91	6585	323.5	1249.8	17	98	8104
Bulgaria–Romania	883.5	2009.4	27	329	12574	553.7	1117.3	9	210	6967
Latin America	1852.1	7011.5	24	411	45416	1780.4	6205.6	62	484	40249
Morocco	285.6	662.5	5	127	4211	303.3	599.2	6	120	3643
China	180.5	751.3	0	42	4846	241.5	876.5	0	51	5629
B: *Migrant stock*, by group-location combination[a]										
EU25	1.32	0.75	0.38	1.13	3.41	1.78	0.79	0.56	1.75	3.63
Bulgaria–Romania	2.44	2.85	0.29	1.37	13.37	5.62	4.81	0.97	3.81	21.07
Latin America	4.56	2.07	1.98	3.93	10.16	4.99	2.06	1.72	4.60	9.23
Morocco	1.47	1.22	0.17	1.09	5.35	1.56	1.20	0.24	1.15	4.95
China	0.14	0.10	0.00	0.14	0.43	0.28	0.21	0.00	0.22	0.95
C: *Location-specific characteristics*										
Population density (1,000)	1.525	1.630	0.132	0.695	6.862	1.684	1.780	0.165	0.962	7.487
GDP per capita (1,000 Euros)	27.982	17.382	9.228	21.575	98.506	27.345	15.229	10.151	22.080	74.589
Employment (number of jobs) per capita[b]	0.336	0.174	0.099	0.298	1.561	0.330	0.180	0.119	0.280	1.513
Avg. annual employment growth (%)	4.55	1.98	0.10	4.69	10.57	3.04	2.36	–3.29	2.70	8.64
Unemployment rate (%)	3.21	0.70	2.04	3.13	5.15	4.13	1.18	2.24	4.22	6.69
Number public transportation lines	20.8	33.4	2	14	221	24.2	37.8	3	17	252
Centrality index	274.1	69.6	123.8	282.2	499.5	295.1	72.8	134.5	303.0	528.3
Gross disp. income per capita (1,000 Euros)	17.359	4.535	11.672	15.750	28.655	18.149	5.034	12.232	16.378	31.826

[a] The migrant stocks are defined as the percentage shares of immigrant communities in the total population in a location.

[b] Employment (number of jobs) per capita is not the same as "employment rate", where the latter would act as a measure of labor-market participation.

Table 3
Econometric results for alternative model specifications.

	Without location-specific effects		With location-specific effects	
			Random	Fixed
	(1)	(2)	(3)	(4)
Dependent variable: Number of migrant arrivals, by group-location combination				
Log population density ($\hat{\beta}_1$)	0.312^a	0.320^a	2.000^a	3.264^a
	(0.093)	(0.084)	(0.447)	(0.557)
Log GDP per capita ($\hat{\beta}_2$)	0.006	–0.009	–0.064	0.303
	(0.343)	(0.348)	(0.423)	(0.346)
Unemployment rate ($\hat{\beta}_3$)	0.153	0.159	–0.073	-0.092^c
	(0.108)	(0.117)	(0.056)	(0.052)
Log employment per capita ($\hat{\delta}_0$)	0.039	0.002	0.805^c	1.293^a
	(0.409)	(0.411)	(0.490)	(0.479)
Log emp. p.c. × Avg. annual employment-growth rate ($\hat{\delta}_1$)	-0.126^a	-0.121^b	0.142^a	0.201^a
	(0.046)	(0.047)	(0.031)	(0.033)
Log emp. p.c. × Avg. annual employment-growth rate squared ($\hat{\delta}_2$)	0.014^b	0.013^b	-0.014^a	-0.018^a
	(0.006)	(0.006)	(0.003)	(0.003)
Log gross disposable income per capita ($\hat{\beta}_4$)	-7.300^c	–6.119	26.315^a	39.538^a
	(4.181)	(3.886)	(7.968)	(8.623)
Log gross disposable income per capita squared ($\hat{\beta}_5$)	1.184^c	0.998	-3.969^a	-5.908^a
	(0.714)	(0.660)	(1.228)	(1.272)
Log number of public transportation lines ($\hat{\beta}_6$)	0.027	–0.047	0.615	1.008^a
	(0.224)	(0.211)	(0.449)	(0.361)
Log number of public transportation lines squared ($\hat{\beta}_7$)	0.145^a	0.157^a	–0.133	-0.230^a
	(0.031)	(0.029)	(0.084)	(0.066)
Log centrality index ($\hat{\beta}_8$)	0.073	0.093	-4.670^b	-7.639^a
	(0.251)	(0.238)	(1.995)	(2.225)
EU25 dummy × Local migrant stock EU25 ($\hat{\mu}_1$)	–0.164	–0.110	–0.198	–0.111
	(0.101)	(0.111)	(0.385)	(0.325)
Bul.–Rom. dummy × Local migrant stock Bul.–Rom. ($\hat{\mu}_2$)	0.126^a	0.155^a	–0.022	-0.025^b
	(0.019)	(0.023)	(0.028)	(0.012)
Lat. Am. dummy × Local migrant stock Lat. Am. ($\hat{\mu}_3$)	0.217^a	0.204^a	0.013	–0.016
	(0.023)	(0.022)	(0.032)	(0.026)
Morocco dummy × Local migrant stock Morocco ($\hat{\mu}_4$)	0.046	0.042	0.183	0.179^c
	(0.072)	(0.075)	(0.112)	(0.095)
China dummy × Local migrant stock China ($\hat{\mu}_5$)	-1.414^b	-1.293^b	1.380	1.283^a
	(0.620)	(0.590)	(1.378)	(0.414)
EU25 × Madrid × Year dummy ($\hat{\theta}_{1Mt}$)	0.618^a	0.738^a	0.422^a	0.463^a
	(0.166)	(0.192)	(0.104)	(0.095)
Bulgaria–Romania × Madrid × Year dummy ($\hat{\theta}_{2Mt}$)	–0.094	0.266	–0.134	–0.093
	(0.246)	(0.232)	(0.103)	(0.092)
Latin America × Madrid × Year dummy ($\hat{\theta}_{3Mt}$)	–0.084	-0.300^b	–0.095	–0.050
	(0.123)	(0.129)	(0.066)	(0.067)
Morocco × Madrid × Year dummy ($\hat{\theta}_{4Mt}$)	-0.515^b	-0.585^a	-0.160^b	–0.104
	(0.239)	(0.195)	(0.063)	(0.065)
China × Madrid × Year dummy ($\hat{\theta}_{5Mt}$)	0.999^a	1.043^a	-0.350^c	-0.294^a
	(0.362)	(0.342)	(0.199)	(0.091)
Group-year dummies (α_{it})	No	Yes	Yes	Yes
Group-location fixed effects (γ_{ij})	No	No	No	Yes
Number of observations	410	410	410	408
Log-likelihood	–25569.2	–23687.2	–2961.7	–1308.2
Likelihood-Ratio chi2 test			41451^a	3307.0^a
(*p*-value)			(0.000)	(0.000)
Hausman chi2 test				274.7^a
(*p*-value)				(0.000)

Table 3
Continued.

	Without individual location effects		With individual location effects	
			Random	Fixed
	(1)	(2)	(3)	(4)
Dependent variable: Number of migrant arrivals, by group-location combination				
Wald chi2 tests				
Group-year dummies jointly = 0		32.06[a]	3.09[b]	36.08[a]
(*p*-value)		(0.000)	(0.010)	(0.000)
Group-Madrid-year dummies jointly = 0	146.1[a]	85.26[a]	7.69[a]	65.54[a]
(*p*-value)	(0.000)	(0.000)	(0.000)	(0.000)
Local migrant stocks jointly = 0	521.6[a]	679.0[a]	0.97	20.40[a]
(*p*-value)	(0.000)	(0.000)	(0.434)	(0.001)

The results in column 4 are for the baseline Poisson specification in Eq. (7). The (maximum) number of observations is equal to the number of origins times the number of destinations times the number of years; that is, $5 \times 41 \times 2 = 410$. Heteroskedasticity-robust standard errors for the pooled models (columns 1–2), jackknife standard errors for the random-effects (RE) model (column 3), and heteroskedasticity-and-clustering robust (HAC) standard errors for the fixed-effects (FE) model (column 4) are given in parentheses. In the FE regression, one panel unit (two observations, for Chinese immigrants) was dropped because of all zero outcomes. The pooled models (without location effects) have been estimated using the `poisson` command in STATA, while the FE and RE models have been estimated using the `xtpoisson` command in STATA. The pooled model in column 1 includes a common year dummy; the common time effect was estimated at −0.476 (significant at the 1% level).

[a] Significant at the 1% level; [b] Significant at the 5% level; [c] Significant at the 10% level.

Table 4
Testing the Poisson variance assumption, by immigrant group.

	A: Gamma GLM, based on raw residuals $\hat{r}_{ij}$		B: OLS, based on Pearson residuals $\tilde{r}_{ij}$	
	$H_0: \lambda_0 = e^{\ln(\lambda_0)} = 1$	$H_0: \lambda_1 = 1$	$H_0: \lambda_0 = 1$	$H_0: \lambda_0(\lambda_1 - 1) = 0$
EU25	2.212	1.096	1.815	0.276
	(0.516)	(0.622)	(0.728)	(0.625)
			2.249	0.254
			(0.609)	(0.678)
Bulgaria–Romania	1.013	1.325[c]	–2.785	1.156
	(0.989)	(0.060)	(0.375)	(0.160)
			–3.045	1.729
			(0.498)	(0.175)
Latin America	6.980	0.903	4.661	–0.119
	(0.537)	(0.688)	(0.286)	(0.847)
			5.474	–0.265
			(0.211)	(0.666)
Morocco	3.531	0.882	3.900[b]	–0.324
	(0.236)	(0.418)	(0.049)	(0.257)
			2.967[c]	–0.196
			(0.057)	(0.362)
China	1.562	1.027	1.835	0.191
	(0.446)	(0.808)	(0.270)	(0.371)
			1.611	0.032
			(0.328)	(0.817)

The p-values, shown in parentheses, are for testing the null hypotheses indicated in the corresponding column heading. The `nlcom` command in STATA (delta method) was used to test the null for $\lambda_0 = e^{\ln(\lambda_0)}$ in panel A of the table. In panel B of the table, the first (second) line refers to the 2005 (2009) Pearson residuals. Key references for these tests are Manning and Mullahy (2001) and Santos Silva and Tenreyo (2006).

[a] Significant at the 1% level; [b] Significant at the 5% level; [c] Significant at the 10% level.

Table 5

Testing spatial autocorrelation of residuals—Moran's I test, by immigrant group.

	EU25	Bulgaria–Romania	Latin America	Morocco	China
A: Without location fixed effects (model in column 2 of Table 3)					
Raw residuals					
I_{RR} for $\hat{r}_{ij,2005}$	0.349[a]	0.041	0.269[a]	–0.003	–0.051
(p-value)	(0.000)	(0.435)	(0.004)	(0.563)	(0.793)
I_{RR} for $\hat{r}_{ij,2009}$	0.321[a]	0.021	–0.017	0.306[a]	0.270[a]
(p-value)	(0.001)	(0.438)	(0.933)	(0.001)	(0.001)
Pearson residuals					
I_{PR} for $\tilde{r}_{ij,2005}$	0.318[a]	0.161[c]	0.302[a]	0.241[a]	–0.088
(p-value)	(0.002)	(0.077)	(0.002)	(0.008)	(0.548)
I_{PR} for $\tilde{r}_{ij,2009}$	0.319[a]	0.030	–0.024	0.339[a]	0.380[a]
(p-value)	(0.001)	(0.568)	(0.995)	(0.001)	(0.000)
B: With location fixed effects (model in column 4 of Table 3)					
Raw residuals					
I_{RR} for $\hat{r}_{ij,2009} = -\hat{r}_{ij,2005}$	0.071	0.099	–0.019	–0.098	–0.033
(p-value)	(0.357)	(0.222)	(0.955)	(0.484)	(0.936)
Pearson residuals					
I_{PR} for $\tilde{r}_{ij,2005}$	0.050	–0.097	0.067	–0.155	0.029
(p-value)	(0.490)	(0.504)	(0.393)	(0.234)	(0.619)
I_{PR} for $\tilde{r}_{ij,2009}$	0.031	–0.080	0.042	–0.157	0.091
(p-value)	(0.604)	(0.612)	(0.535)	(0.227)	(0.287)

In contrast to the raw residuals, $\tilde{r}_{ij,2009} \neq -\tilde{r}_{ij,2005}$ in general. The mean and variance of Moran's I statistic are determined under the normality assumption. Based on the Shapiro-Wilk W test for normality, we found that normality of the Pearson residuals could not be rejected for the baseline model (with individual location effects). See also Lin and Zhang (2007).

[a] Significant at the 1% level; [b] Significant at the 5% level; [c] Significant at the 10% level.

Table 6

Estimated expected changes in Chinese immigration, for selected municipalities.

	Initial stock of Chinese immigrants in 2008		One thousand added to the migrant stock[a]	Expected change in choice probability ($\Delta\hat{p}$ in % points)[b]		Expected change in number of new immigrants[b]	
	(number)	(% of population)	(Δx in % points)	CL	Poisson	CL	Poisson
Madrid	23,726	0.74144	0.03125	0.76[c]	1.76[c]	75[c]	174[c]
Parla	1,031	0.95418	0.92549	7.95	8.57	788	849
Fuenlabrada	1,055	0.54161	0.51337	3.22	3.39	319	336
Tres Cantos	106	0.26105	2.46269	1.63	1.64	163	164
			(Δx = 0.1% point)				
Madrid			3,200	2.43[c]	5.62[c]	240[c]	556[c]
Parla			108	0.86	0.93	85	92
Fuenlabrada			195	0.63	0.66	62	65
Tres Cantos			41	0.07	0.07	7	7

[a] Holding population size constant (recall that the migrant stock has been measured as the percentage share of the prior immigrant population in the total population in each municipality within the MMA.

[b] The change in choice probability for the CL interpretation of the Poisson estimate is $100[\partial\,\hat{p}/\partial x_j]\cdot\Delta x_j = 100\hat{\beta}_j\hat{p}(1-\hat{p})\cdot\Delta x_j = 100\hat{\beta}_j\hat{p}(1-\hat{p})\cdot\Delta x_j$, where $\hat{p}$ is the predicted 2009 proportion of Chinese immigrants to the location under study in the total number of Chinese immigrants to the MMA (Schmidheiny and Brülhart, 2012, p. 1089). The change in choice probability for the Poisson model is $100[\partial\,\hat{p}/\partial x_j]\cdot\Delta x_j = 100\hat{\beta}_j\hat{p}\cdot\Delta x_j = 100\hat{\beta}_j\hat{p}\cdot\Delta x_j$. The marginal effect for Poisson (using general notation) is $[\partial\widehat{E(y|\mathbf{x})}/\partial x_j]\cdot\Delta x_j = \hat{\beta}_j\exp(\mathbf{x}'\hat{\boldsymbol{\beta}})\cdot\Delta x_j = \hat{\beta}_j\hat{y}\cdot\Delta x_j$ (Wooldridge, 2002, p. 648). The CL interpretation of this marginal effect is $[\partial\widehat{E(y|\mathbf{x})}/\partial x_j](1-\hat{p})\cdot\Delta x_j = \hat{\beta}_j\exp(\mathbf{x}'\hat{\boldsymbol{\beta}})(1-\hat{p})\cdot\Delta x_j = \hat{\beta}_j\hat{y}(1-\hat{p})\cdot\Delta x_j$. The Poisson estimate is $\hat{\beta}_j = 1.283$ (see column 4 of Table 3). The 2008 population sizes are: Madrid 3,200,000; Parla 108,051; Fuenlabrada 194,791; Tres Cantos 40,606. The 2009 numbers of new Chinese migrant arrivals ($\hat{y}$) predicted by our baseline Poisson regression are: Madrid 5,626; Parla 715; Fuenlabrada 510; Tres Cantos 52. The 2009 proportions of Chinese immigrants ($\hat{p}$) predicted by the Poisson model are: Madrid 0.5685; Parla 0.0722; Fuenlabrada 0.0515; Tres Cantos 0.0052.

[c] The expected changes for the city of Madrid have been corrected for the individual "Madrid effect", where θ_{5Mt} was estimated at −0.294.

Fig. 1. Spatial distribution of 2009 immigration rates—Location Quotients.

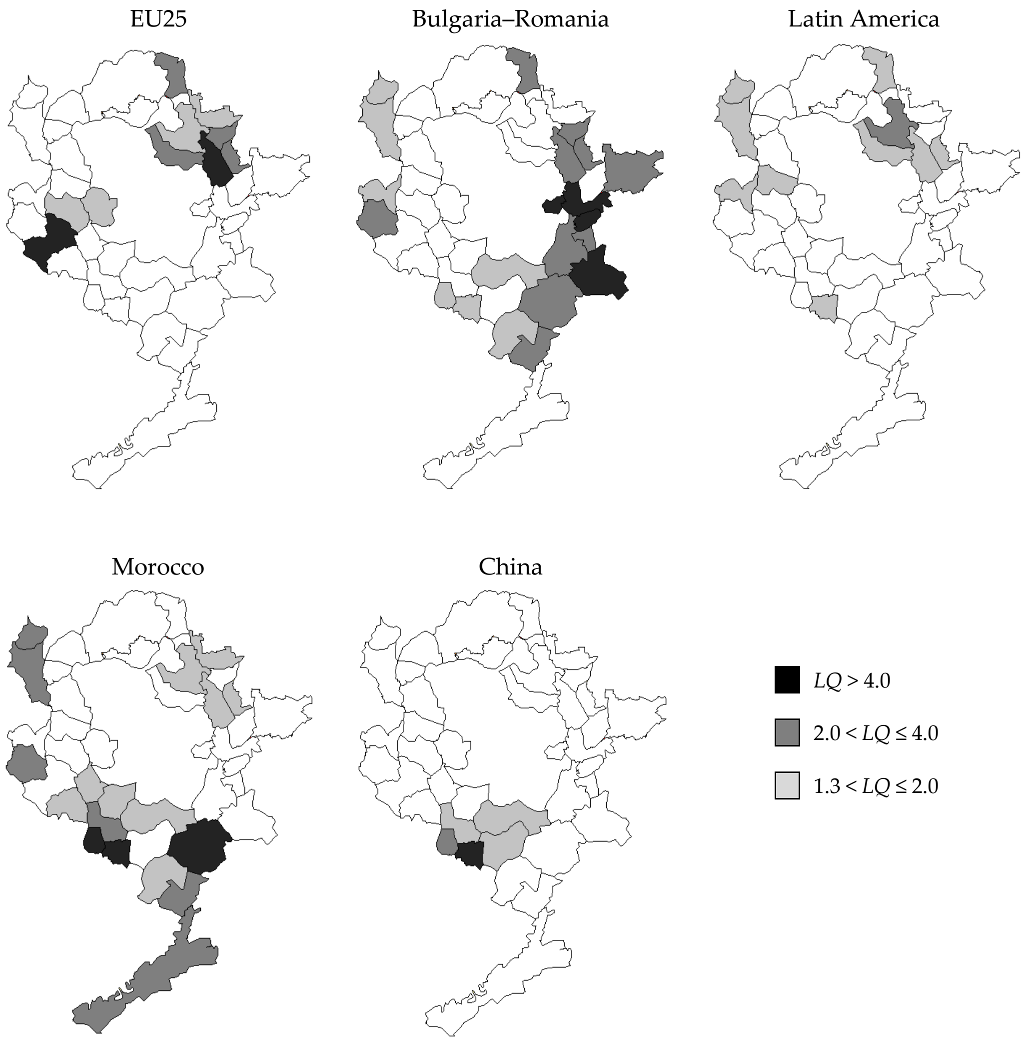
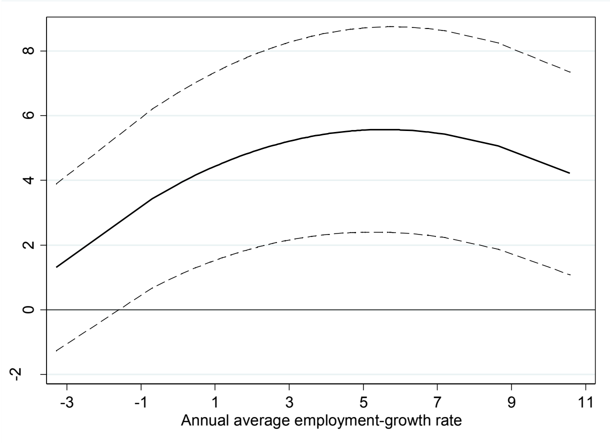
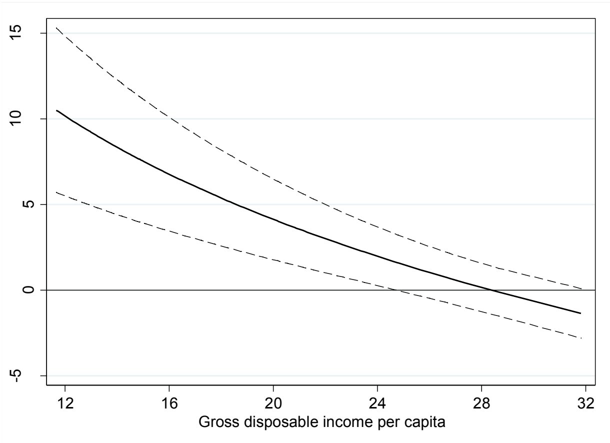


Fig. 2. Estimated partial effects for non-linear (quadratic) relationships.

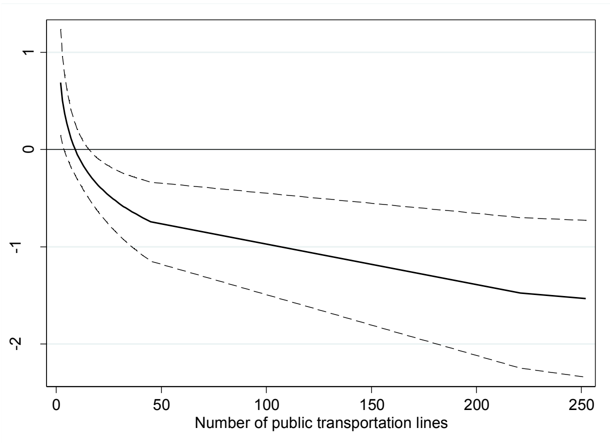
A: Employment (number of jobs) per capita



B: Gross disposable income per capita



C: Number of public transportation lines



The solid curves represent the estimated partial effects (percentage changes) evaluated at different values of variable on the horizontal axis. The dashed curves mark the upper and lower bounds of the 95% confidence interval. The partial effect in panel A is evaluated at the sample mean of employment (jobs) per capita.

Appendix A: Data sources and definition of variables

A.1. Immigration data

Immigration data (2005, 2009) are taken from the Spanish National Statistical Institute (http://www.ine.es), Estadística de Variaciones Residenciales (EVR). These data cover documented and undocumented immigrants (individuals, not households) to all municipalities in Spain, by country of nationality, age, and gender. The data are collected by exploiting the data from the local *padrones* on new registrations and deletions due to changes in the municipality of residence.

A.2. Migrant stocks

Data on migrant stocks (2004, 2008) are collected from the Statistical Institute of the Comunidad de Madrid, Padrón continuo. Migrant stocks for the five distinct immigrant populations, expressed as percentages of the local population. Since both legal and illegal immigrant have been enumerated, the *padrón* data are likely to provide a relatively accurate head-count of the actual number of established and new immigrants by municipality.

A.3. Economic and demographic variables

Data on population, GDP per capita, and gross disposable income per capita (2004, 2008) are drawn from the Statistical Institute of the Comunidad de Madrid. Data on unemployment rates are from the Servicio Público de Empleo Estatal (SEPE), Ministerio de Empleo y Seguridad Social. Data on employment (number of jobs) per capita and employment-growth rates are from the Colectivo Empresarial de la Comunidad de Madrid, Instituto de Estadística de la Comunidad de Madrid.

Employment-growth rates have been calculated as 3-year average annual growth rates (expressed as percentages), covering the four years prior to immigration (2001–2004, 2005–2008). The employment level for 2001 has been estimated from the provincial and sub-regional units published by the Encuesta de Población Activa (INE) and Instituto de Estadística de la Comunidad de Madrid, respectively.

The levels of employment (number of jobs) per capita and unemployment rates (expressed as percentages) are for 2004 and 2008.

A.4 Accessibility and other spatial measures

Data on the number of public transportation lines (2004, 2008) are from the Consorcio Regional de Transportes de Madrid, Instituto de Estadística de la Comunidad de Madrid. Data on area (in square kilometers) are from the Consejería de Medio Ambiente y Ordenación del Territorio, Comunidad de Madrid. Distance and contiguity matrixes have been constructed at the L.R. Klein Institute, Autonomous University of Madrid.

A.5. Two-period panel

We use a two-year ($t = 2005, 2009$) panel of 41 destinations ($J = 41$) and five origin groups ($I = 5$), providing $T \times J \times I = 2 \times 41 \times 5 = 410$ (potential) observations. The five groups of immigrants considered are: EU25 countries (excluding Spanish nationals), Bulgaria–Romania, Latin America (Spanish speaking immigrants only), Morocco, and China. For a list of the origin countries by area, see Appendix C, Table C.3.

<u>**Appendix B: Raw residuals for two-period conditional fixed-effects Poisson model**</u>

For some non-linear models, the fixed effects can be removed from the likelihood function by *conditioning* on a sufficient statistic. The fixed-effects Poisson regression conditions on the *total counts* within each panel unit—here, each group-destination (i, j) combination.

Starting from this property, Baltagi (2009, p. 230) has shown that the fixed effects can be recovered as follows (omitting $\hat{\alpha}_{it}$, etc., for clarity):

$$\hat{\gamma}_{ij} = \ln\left[\frac{\sum_{t=1}^{T} n_{ijt}}{\sum_{t=1}^{T} e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}}}\right] \tag{B.1}$$

from which the raw residuals can be derived as $\hat{r}_{ijt} = n_{ijt} - e^{\hat{\gamma}_{ij} + \hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}}$. With only two time periods $(t = 1, 2)$, it can be shown that the "within"-transformation (demeaning) implies that $\hat{r}_{ij,1} = -\hat{r}_{ij,2}$. Specifically, given that

$$\begin{aligned} \hat{r}_{ijt} &= n_{ijt} - e^{\hat{\gamma}_{ij} + \hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}} \\ &= n_{ijt} - e^{\hat{\gamma}_{ij}}\, e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}} \\ &= n_{ijt} - \exp\left\{\ln\left[\frac{\sum_{t=1}^{T} n_{ijt}}{\sum_{t=1}^{T} e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}}}\right]\right\} e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}} \\ &= n_{ijt} - \left[\frac{\sum_{t=1}^{T} n_{ijt}}{\sum_{t=1}^{T} e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}}}\right] e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{jt}} \end{aligned} \tag{B.2}$$

it follows that

$$\begin{aligned} \hat{r}_{ij,1} &= n_{ij,1} - \left[\frac{n_{ij,1} + n_{ij,2}}{e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}}\right] e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} \\ &= \frac{n_{ij,1}(e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}) - (n_{ij,1} + n_{ij,2})e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}}}{e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}} \\ &= \frac{n_{ij,1}e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}} - n_{ij,2}e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}}}{e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}} \end{aligned} \tag{B.3}$$

and

$$\begin{aligned} \hat{r}_{ij,2} &= n_{ij,2} - \left[\frac{n_{ij,1} + n_{ij,2}}{e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}}\right] e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}} \\ &= \frac{n_{ij,2}(e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}) - (n_{ij,1} + n_{ij,2})e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}}{e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}} \\ &= \frac{-(n_{ij,1}e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}} - n_{ij,2}e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}})}{e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,1}} + e^{\hat{\boldsymbol{\beta}}'\mathbf{y}_{j,2}}} \\ &= -\hat{r}_{ij,1} \end{aligned} \tag{B.4}$$

where it should be noted that this relationship does not generally hold for the Pearson residuals, $\tilde{r}_{ijt} = \hat{r}_{ijt}/\sqrt{\hat{n}_{ijt}}$.

Appendix C: Tables and figures

Table C.1
Results of exploratory shift-share analysis.

	2005	2009	Difference	Share	Area shift	Group shift
A: *City of Madrid*						
EU25	6585	8104	1519	−564	1013	1070
Bulgaria–Romania	12574	6967	−5607	−1077	−913	−3617
Latin America	45416	40249	−5167	−3889	−3407	2129
Morocco	4211	3643	−568	−361	−828	621
China	4846	5629	783	−415	−855	2053
Total	73632	64592	−9040	−6305	−4989	2255
			(73.5%)			
B. *Rest of metro area*						
EU25	5731	5158	−573	−491	−1013	931
Bulgaria–Romania	23649	15733	−7916	−2025	913	−6804
Latin America	30522	32746	2224	−2614	3407	1431
Morocco	7500	8791	1291	−642	828	1105
China	2553	4271	1718	−219	855	1082
Total	69955	66699	−3256	−5991	4989	−2255
			(26.5%)			
C: *Total metro area*						
EU25	12316	13262	946	−1055	0	2001
Bulgaria–Romania	36223	22700	−13523	−3102	0	−10421
Latin America	75938	72995	−2943	−6503	0	3560
Morocco	11711	12434	723	−1003	0	1726
China	7399	9900	2501	−634	0	3135
Total	143587	131291	−12296	−12296	0	0
			(100.0%)			

Notation: n is the number of international migrants, areas $a = 1,2$ (city of Madrid and rest of metropolitan area, respectively), and groups $i = 1, \dots, 5$ (EU25, Bulgaria–Romania, Latin America, Morocco, and China). The shift-share decomposition of differences in immigration numbers is as follows:

$$n_a^{09} - n_a^{05} = \underbrace{\sum_{i=1}^{5} n_{i,a}^{05} \left(\frac{n^{09}}{n^{05}} - 1 \right)}_{\text{Share}} + \underbrace{\sum_{i=1}^{5} n_{i,a}^{05} \left(\frac{n_{i,a}^{09}}{n_{i,a}^{05}} - \frac{n_i^{09}}{n_i^{05}} \right)}_{\text{Area shift}} + \underbrace{\sum_{i=1}^{5} n_{i,a}^{05} \left(\frac{n_i^{09}}{n_i^{05}} - \frac{n^{09}}{n^{05}} \right)}_{\text{Group shift}}, \quad a = 1,2$$

Table C.2
Results of sign test for commonalities of “within”-variations, 2005–2009.

	Positive	Negative	Zero	p-Value
Population density (PD_j)	41	0	0	0.000
GDP per capita (GDP_j^{pc})	20	21	0	1.000
Average annual employment-growth rate (ΔEMP_j)	11	30	0	0.004
Employment per capita (EMP_j^{pc})	16	25	0	0.211
Unemployment rate (UR_j)	38	3	0	0.000
Gross disposable income per capita (INC_j^{pc})	39	2	0	0.000
Number of public transportation lines (PTL_j)	33	4	4	0.000
Centrality index (CI_j)	41	0	0	0.000
Local migrant stocks:				
EU25 (MS_{1j})	40	1	0	0.000
Bulgaria–Romania (MS_{2j})	41	0	0	0.000
Latin America (MS_{3j})	27	14	0	0.060
Morocco (MS_{4j})	27	14	0	0.060
China (MS_{5j})	35	2	4	0.000

The sign test (Snedecor and Cochran, 1989) is implemented using the `signtest` command in STATA. The p-values are for two-sided tests.

Table C.3
List of origin countries (immigrant groups and their compositions).

EU25	%2005	%2009	Latin America	%2005	%2009	Eastern Europe	%2005	%2009
Austria	0.6%	0.9%	Argentina	5.5%	3.7%	Bulgaria	11.5%	11.4%
Belgium	1.6%	1.5%	Bolivia	16.0%	7.0%	Romania	88.5%	88.6%
Cyprus	0.0%	0.1%	Chili	2.7%	2.0%	Total	100.0%	100.0%
Czech Republic	0.5%	0.8%	Columbia	15.6%	19.3%			
Denmark	0.7%	0.6%	Costa Rica	0.1%	0.2%	**Africa**		
Estonia	0.1%	0.2%	Cuba	2.3%	2.8%	Morocco	100.0%	100.0%
Finland	0.5%	0.5%	Dominican Republic	8.4%	8.4%			
France	12.4%	16.8%	Ecuador	17.8%	16.1%	**Asia**		
Germany	8.2%	8.5%	El Salvador	0.3%	0.6%	China	100.0%	100.0%
Greece	1.6%	1.1%	Guatemala	0.3%	0.5%			
Hungary	0.7%	0.8%	Honduras	0.6%	1.8%			
Ireland	1.0%	1.2%	Mexico	2.1%	2.1%			
Italy	22.1%	26.8%	Nicaragua	0.2%	0.8%			
Latvia	0.2%	0.5%	Panamá	0.2%	0.2%			
Lithuania	0.8%	0.8%	Paraguay	6.3%	10.0%			
Luxembourg	0.0%	0.0%	Peru	16.4%	15.6%			
Malta	0.0%	0.0%	Uruguay	0.9%	0.6%			
Netherlands	3.4%	3.2%	Venezuela	4.5%	4.4%			
Poland	24.7%	12.0%	Total	100.0%	100.0%			
Portugal	12.1%	12.4%						
Slovakia	0.6%	0.5%						
Slovenia	0.2%	0.1%						
Sweden	1.4%	1.7%						
United Kingdom	6.5%	9.1%						
Total	100.0%	100.0%						

Table C.4
Effect of initial (2005) immigrations on subsequent (2009) immigrations.

Dependent variable: Number of migrant arrivals, by group-location combination	
Log initial (2005) immigration numbers ($\hat{\beta}_0$)	0.748[a]
	(0.050)
Log population density ($\hat{\beta}_1$)	0.037
	(0.057)
Log GDP per capita ($\hat{\beta}_2$)	–0.119
	(0.234)
Unemployment rate ($\hat{\beta}_3$)	0.080
	(0.066)
Log employment per capita ($\hat{\delta}_0$)	0.311
	(0.255)
Log emp. p.c. × Avg. annual employment-growth rate ($\hat{\delta}_1$)	–0.002
	(0.023)
Log emp. p.c. × Avg. annual employment-growth rate squared. ($\hat{\delta}_2$)	–0.004
	(0.003)
Log gross disposable income per capita ($\hat{\beta}_4$)	7.199[a]
	(2.382)
Log gross disposable income per capita squared ($\hat{\beta}_5$)	–1.244[a]
	(0.389)
Log number of public transportation lines ($\hat{\beta}_6$)	0.579[b]
	(0.285)
Log number of public transportation lines squared ($\hat{\beta}_7$)	–0.075
	(0.052)
Log centrality index ($\hat{\beta}_8$)	0.236[c]
	(0.140)
EU25 dummy × Local migrant stock EU25 ($\hat{\mu}_1$)	0.160[b]
	(0.066)
Bulgaria–Romania dummy × Local migrant stock Bulgaria–Romania ($\hat{\mu}_2$)	0.016
	(0.011)
Latin America dummy × Local migrant stock Latin America ($\hat{\mu}_3$)	0.031[c]
	(0.018)
Morocco dummy × Local migrant stock Morocco ($\hat{\mu}_4$)	0.127[a]
	(0.043)
China dummy × Local migrant stock China ($\hat{\mu}_5$)	1.428[a]
	(0.238)
EU25 × Madrid dummy ($\hat{\theta}_{1M}$)	1.175[a]
	(0.372)
Bulgaria–Romania dummy ($\hat{\theta}_{2M}$)	0.562
	(0.376)
Latin America × Madrid dummy ($\hat{\theta}_{3M}$)	0.619[c]
	(0.351)
Morocco × Madrid dummy ($\hat{\theta}_{4M}$)	0.521
	(0.375)
China × Madrid dummy ($\hat{\theta}_{5M}$)	0.205
	(0.379)
Group dummies (α_i)	Yes
Number of observations	201
Log-likelihood	–2132.8
Wald chi2 test	
Group dummies jointly = 0	33.20[a]
(*p*-value)	(0.000)
Group-Madrid dummies jointly = 0	135.50[a]
(*p*-value)	(0.000)
Local migrant stocks jointly = 0	44.49[a]
(*p*-value)	(0.000)

[a] Significant at the 1% level; [b] Significant at the 5% level; [c] Significant at the 10% level.

Fig. C.1. Base map of Madrid metropolitan area.