I Z A

IZA DP No. 5958

Multilateral Resistance to Migration

Simone Bertoli
Jesús Fernández-Huertas Moraga

September 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Multilateral Resistance to Migration

Simone Bertoli
*CERDI, University of Auvergne
and CNRS*

Jesús Fernández-Huertas Moraga
*FEDEA, IAE (CSIC)
and IZA*

ABSTRACT

Multilateral Resistance to Migration[*]

The rate of migration observed between two countries does not depend solely on their relative attractiveness, but also on the one of alternative destinations. Following the trade literature, we term the influence exerted by other destinations on bilateral flows as Multilateral Resistance to Migration, and we show how it can be accounted for when estimating the determinants of migration flows in the context of a general individual random utility maximization model. We propose the use of the Common Correlated Effects estimator (Pesaran, 2006) and apply it to high-frequency data on the Spanish immigration boom between 1997 and 2009. Compared to more restrictive estimation strategies developed in the literature, the bias goes in the expected direction: we find a smaller effect of GDP per capita and a larger effect of migration policies on bilateral flows.

Corresponding author:

Jesús Fernández-Huertas Moraga
FEDEA
Jorge Juan, 46
28001 Madrid
Spain
E-mail: jfernandezhuertas@fedea.es

1   Introduction

The responsiveness of the scale of migration flows to varying economic conditions - both in sending and recipient countries - and to changing immigration policies at destination represents a central topic in the international migration literature. While some recent contributions have provided econometric analysis of aggregate data where the identification strategy is consistent with the proposed underlying individual-level migration decision model (Beine, Docquier, and Ozden, 2011; Grogger and Hanson, 2011; Ortega and Peri, 2009),[1] others have relied on econometric specifications that have not been fully micro-founded (Clark, Hatton, and Williamson, 2007; Pedersen, Pytlikova, and Smith, 2008; Mayda, 2010; Theoharides, McKenzie, and Yang, 2010).

This methodological difference notwithstanding, these papers share a crucial feature, as Hanson (2010) observes that the literature is characterized by a long-standing tradition of "estimating bilateral migration flows as a function of characteristics in the source and destination countries only". Still, would-be migrants sort themselves across alternative destinations, so that it is important to understand whether this econometric approach allows to control for the possible dependence of the migration rate between any pair of countries upon the time-varying attractiveness of other migrants' destinations. Hanson (2010) argues that "failing to control other migration opportunities could [...] produce biased estimates", and this issue resembles the one raised by Anderson and van Wincoop (2004) with respect to the estimation of the determinants of bilateral trade flows.

Trade between two countries does not depend on bilateral trade costs only, but rather on the relationship between these costs and the costs with the other trading partners; Anderson and van Wincoop (2004) refer to the attractiveness of trading with other partners as multilateral resistance to trade.[2] Similarly, migration flows between a dyad represented by an origin and a destination country do not depend solely on the attractiveness of the latter, but also on how this relates to the opportunities to move to other destinations. Following the terminology introduced by Anderson and van Wincoop (2004), we refer to the attractiveness of other destinations as Multilateral Resistance to Migration.

---

[1]Bertoli, Fernández-Huertas Moraga, and Ortega (2010) analyze the income-sensitivity of international migration flows using individual-level data.

[2]Baldwin (2006) observes that this is nothing more than a specific case of the general principle that "relative prices matter".

This paper directly addresses the concern raised by Hanson (2010). First, it relates the stochastic properties of the underlying individual migration decision model to the need to control for multilateral resistance to migration when estimating the determinants of bilateral migration flows. Second, it shows that the data usually employed in the literature suffice to obtain consistent estimates even when multilateral resistance to migration matters. Third, it applies the proposed econometric approach - which draws on Pesaran (2006) - to analyze the determinants of migration flows to Spain over 1997-2009 using high-frequency administrative data.

The paper presents a general random utility maximization (RUM) model that describes the migration decision problem that individuals face. The theoretical model shows that multilateral resistance to migration represents an issue for the analysis of aggregate data whenever the stochastic component of location-specific utility is such that the independence of irrelevant alternatives assumption fails.[3] The derivation of the econometric specification from the random utility maximization model reveals that multilateral resistance to migration, which is unobservable for the econometrician, gives rise to an endogeneity problem, as the regressors are correlated with the error term, which also exhibits serial and spatial correlation.

We show that the multilateral resistance to migration term entering the error of the equation that describes the determinants of aggregate migration flows on the basis of the RUM model can be expressed as the inner product of a vector of dyad-specific factor loadings and a vector of time-specific common effects. This entails that the structure of the error term coincides with the multifactor error model presented in Pesaran (2006). Pesaran (2006) proposed an estimator, the Common Correlated Effects (CCE) estimator, which allows to derive consistent estimates from panel data when the error follows this structure, i.e. it is serially and spatially correlated, and the regressors are endogenous.[4] The CCE estimator requires to estimate a regression where the cross-sectional averages of the dependent and of all the independent variables are included as auxiliary regressors: consistency of the estimates follows from the fact that the multilateral resistance to migration term can be approximated

---

[3]The converse is also true: if the independence of irrelevant alternatives characterizes the individual migration decision problem, then the time-varying attractiveness of other destinations can be disregarded in the econometric analysis, as in Grogger and Hanson (2011) and Beine, Docquier, and Ozden (2011).

[4]Driscoll and Kraay (1998) allow to address the violation of the classical assumptions on the error term, but still require exogeneity of the regressors, which does not hold when multilateral resistance to migration is an issue.

by a dyad-specific linear combination of the cross-sectional averages (Pesaran, 2006).

The adoption of the CCE estimator allows us to address the challenge posed by multilateral resistance to migration using the same data that are traditionally employed in the literature. This approach is more general than the one proposed in Mayda (2010), who includes a weighted average of income per capita in the other destinations as a control for their time-varying attractiveness,[5] and the one in Ortega and Peri (2009), which is valid only under more restrictive assumptions on the underlying RUM model and which does not allow to identify the effects of origin-specific variables.

The proposed econometric approach is applied to the analysis of the determinants of bilateral migration flows to Spain between 1997 and 2009, when this country experienced an unprecedented boom in immigration. In fact, Spain recorded "the highest rate of growth of the foreign-born population over a short period observed in any OECD country since the Second World War" (OECD, 2010): the immigrant share went from 3 percent of the population in 1998 to 14 percent in 2009 (INE, 2010b).[6] Migration data come from the *Estadística de Variaciones Residenciales* (EVR; (INE, 2010a)), an administrative dataset collected by the *Instituto Nacional de Estadística*. A key feature of the EVR is that it provides us with high-frequency data, which give to the dataset the longitudinal dimension that is required to be confident about the application of the CCE estimator (Pesaran, 2006).

The data from the EVR, which have been aggregated by quarter, have been combined with data from IMF (2010a) and World Bank (2010) on real GDP and population at origin for 61 countries,[7] which represent 87 percent of the total flows to Spain over our period of analysis. Furthermore, we have compiled information about the various facets of Spanish immigration policies - such as bilateral visa waivers and agreements on the portability of pension rights - which have been shown to be relevant determinants of recent immigration to Spain (Bertoli, Fernández-Huertas Moraga, and Ortega, 2011).

Our results show that ignoring the multilateral resistance to migration term biases the estimation of the determinants of migration flows to Spain. In addition, the direction of the

---

[5]Hanson (2010) wonders whether this is "a sufficient statistic for other migration opportunities". We show that this is not the case in general.

[6]These figures can only be compared with Israel in the 1990s, when "immigration increased Israel's population by 12 percent between 1990 and 1994, after emigration restrictions were lifted in an unstable Soviet Union" (Friedberg, 2001), at a time when Israel had not joined the OECD yet.

[7]Data from the International Financial Statistics (IMF, 2010a) have been also combined with data from the World Economic Outlook (IMF, 2010b), and various Central Banks, as described in the Appendix A.3.

bias is the one we could expect. The effect of GDP at origin on migration flows to Spain is two thirds of that found in a specification that does not control for multilateral resistance to migration, although it is still negative and significant: a 1 percent drop in GDP per capita in a country increases its emigration rate to Spain by 3.1 percent. This bias is in the opposite direction of that found on the impact of migration policies. The only migration policy that has a significant effect on migration flows to Spain is the adoption of a visa waiver. This effect only turns significant when the multilateral resistance to migration is accounted for: establishing a visa waiver for a country multiplies its emigration rate to Spain by a factor of 4,[8] while the estimated effect when multilateral resistance to migration is not controlled for is not significantly different from zero.

The paper is related to four strands of the literature. First, the papers that analyze the determinants of bilateral migration flows using panel data in a multi-origin multi-destination framework (Clark, Hatton, and Williamson, 2007; Lewer and den Berg, 2008; Grogger and Hanson, 2011; Mayda, 2010; Ortega and Peri, 2009; Simpson and Sparber, 2010; Pedersen, Pytlikova, and Smith, 2008; Beine, Docquier, and Ozden, 2011). Our theoretical model can also be applied to that framework but, in terms of the structure of the data, our paper is more closely related to Clark, Hatton, and Williamson (2007) and Theoharides, McKenzie, and Yang (2010), which estimate the determinants of bilateral flows to one destination, the United States, and from one origin, the Philippines, respectively.[9]

Second, we draw on the papers that have analyzed high-frequency migration data. Specifically, Hanson and Spilimbergo (1999) and Orrenius and Zavodny (2003) who analyze monthly migration flows from Mexico to the United States.

Third, the theoretical and empirical analysis presented here is related to the papers in the trade literature that discuss the relevance of multilateral resistance to trade (Anderson and van Wincoop, 2003, 2004; Baldwin, 2006).

Fourth, the paper is related to the contributions in the econometric literature that present estimators which allow to deal with violations on the classical assumption about the variance

---

[8]This huge effect is in line with the findings of Bertoli, Fernández-Huertas Moraga, and Ortega (2011) for the case of Ecuadorian migration to Spain.

[9]The analysis is also related to the papers that estimate the influence of demographic factors (Hanson and McIntosh, 2010b,a) and migration networks (Edin, Fredriksson, and Åslund, 2003; Munshi, 2003; McKenzie and Rapoport, 2010; Bertoli, 2010) upon migration flows; these effects are controlled for but not estimated in our paper.

structure of the error term (Driscoll and Kraay, 1998; Hoechle, 2007; Coakley, Fuertes, and Smith, 2002), and with the endogeneity of the regressors (Pesaran, 2006; Bai, 2009; Pesaran and Tosetti, 2011).[10]

The paper is structured as follows: Section 2 presents the random utility maximization model that represents the individual migration decision problem; Section 3 analyzes the relationship between the stochastic properties of the RUM model and the need to control for multilateral resistance to migration in the econometric analysis through the CCE estimator proposed by Pesaran (2006). Section 4 presents the sources of the data used in the econometric analysis and the descriptive statistics. Section 5 discusses the estimates, and the empirical relevance of multilateral resistance to migration for the case that we have analyzed. Finally, Section 6 draws the main conclusions.

2 From individual decisions to aggregate flows

We present here a random utility maximization model that describes the location choice problem that would-be migrants face, which gives us the basis for deriving the determinants of bilateral aggregate migration flows. To keep it as general as possible, we do not specify the factors that influence location-specific utility.

2.1 Random utility maximization model

Consider a set of individuals, indexed by i, originating from a country j belonging to a set H, who have to chose their preferred location among countries belonging to the set $D_j = D \cup \{j\}$, which contains $n(j)$ elements. Let the elements in D_j be indexed by k; the utility that the individual i from country j obtains from opting for country k is given by:

---

[10]Endogeneity of some of the regressors, such as GDP at origin, goes beyond the effect exerted by multilateral resistance to migration: Mishra (2007) and Docquier, Ozden, and Peri (2010) show how wages at origin respond to migration whereas Borjas (2003) and Ottaviano and Peri (2010) among many others show how wages at destination respond to migration, and Bugamelli and Paternó (2009) analyze the relationship between migrants' remittances and current account reversals, and they conclude that remittances lower the probability of such a reversal; Anderson (2011) explores the implications for the estimation strategy when GDP is endogenous to migration flows.

$$U_{ijk} = V_{jk} + \epsilon_{ijk} = \boldsymbol{\beta}' \boldsymbol{x}_{jk} + \epsilon_{ijk} \tag{1}$$

where $\boldsymbol{x}_{jk}$ is a vector of factors - which can include location- or dyad-specific elements,[11] and ϵ_{ijk} is a stochastic term.

The vector $\boldsymbol{p}_{ij} = (p_{ij1}, ..., p_{ijk}, ...)$ which collects the choice probabilities for individual i over all the countries belonging to the set D_j depends on the assumptions about the distribution of the stochastic term in (1). We consider here distributions of ϵ_{ijk} which can be obtained from a Generalized Extreme Value generating function (McFadden, 1978), as the econometric approaches adopted in the literature are all consistent with different GEV models.

Consider a real-valued function G with domain $\mathbb{R}^{n(j)}$,[12] and which takes as its arguments the exponentiated values of the deterministic component in (1), i.e. $Y_{jl} = e^{V_{jl}}$: if G satisfies the four properties described in McFadden (1978),[13] then G is a GEV generating function and the element k in the vector of choice probabilities $\boldsymbol{p}_{ij}$ is equal to the elasticity of G with respect to Y_{jk}.[14]

A simplified version of the GEV generating function proposed by Wen and Koppelman (2001) allows us to present in a unified framework various approaches that have been adopted to estimate the determinants of bilateral migration flows, and the more general approach that we present in this paper. Consider the following GEV generating function:[15]

$$G(Y_{ij1}, ..., Y_{ijn(j)}) = \sum_m \left(\sum_{l \in b_m} (\alpha_{jlm} Y_{jl})^{1/\tau} \right)^\tau \tag{2}$$

where $Y_{jl} = e^{V_{jl}}$ for $l \in D_j$ and b are nests of D_j indexed by m. The matrix $\boldsymbol{\alpha}_j$ collects the allocation parameters α_{jlm}, which characterize the portion of country l which is assigned to the nest b_m for individuals from the origin country j,[16] and τ, with $\tau \in (0, 1]$, is the

---

[11]Location-specific elements vary only over k, while dyad-specific elements vary over each pair (j, k).

[12]Observe that we omit the subscript j from the function G in (2) for the sake of simplicity.

[13]G is nonnegative and homogeneous of degree 1, it diverges to infinity when one its argument diverges to infinity, the partial derivative with respect to any of its argument is nonnegative, and cross-derivatives alternate their signs.

[14]See also Train (2003) for an introduction to GEV models.

[15]Wen and Koppelman (2001) demonstrate that G satisfies the four identifying properties in McFadden (1978).

[16]The allocation parameters satisfy $\alpha_{jlm} \in [0, 1]$ for all $l \in D_j$, and the sum of the elements in each row vector $\boldsymbol{\alpha}_{jl}$ is equal to 1.

dissimilarity parameter for the nests b_m.

The specification in (2) does *not* restrict individuals from different origin countries to have identical preferences, as the allocation matrix $\boldsymbol{\alpha}_j$ can vary across origins. This implies that the stochastic component of utility can follow origin-specific patterns of correlation across alternative destinations.[17]

When the GEV generating function is as in (2), the element k in the vector of choice probabilities $\boldsymbol{p}_{ij}$ is equal to:[18]

$$p_{ijk} = \frac{\sum_m (\alpha_{jkm} Y_{jk})^{1/\tau} \left(\sum_{l \in b_m} (\alpha_{jlm} Y_{jl})^{1/\tau} \right)^{\tau-1}}{\sum_m \left(\sum_{l \in b_m} (\alpha_{jlm} Y_{jl})^{1/\tau} \right)^{\tau}} \tag{3}$$

Wen and Koppelman (2001) refer to the discrete choice model whose choice probabilities are described in (3) as the generalized nested logit model. The relative probability of opting for destination k over staying in the home country j is equal to:

$$\frac{p_{ijk}}{p_{ijj}} = \frac{\sum_m (\alpha_{jkm} Y_{jk})^{1/\tau} \left(\sum_{l \in b_m} (\alpha_{jlm} Y_{jl})^{1/\tau} \right)^{\tau-1}}{\sum_m (\alpha_{jjm} Y_{jj})^{1/\tau} \left(\sum_{l \in b_m} (\alpha_{jlm} Y_{jl})^{1/\tau} \right)^{\tau-1}} \tag{4}$$

If we assume that the origin country j belongs only to a singleton,[19] then we can express the log odds as follows:

$$\ln \left(\frac{p_{ijk}}{p_{ijj}} \right) = \frac{V_{jk}}{\tau} - V_{jj} + \ln \left(\sum_m (\alpha_{jkm})^{1/\tau} \left(\sum_{l \in b_m} (\alpha_{jlm} e^{V_{jl}})^{1/\tau} \right)^{\tau-1} \right) \tag{5}$$

---

[17]When the discrete choice model is generated by (2), ϵ_{ijk} contains additive nest-specific stochastic components; the correlation of the unobserved component of utility in (1) between two different origin-destination dyads depends on (*i*) the allocation vector $\boldsymbol{\alpha}$ corresponding to each dyad, and (*ii*) on the dissimilarity parameter τ, which is inversely related to the correlation across alternatives of the nest-specific stochastic components. If the inner product of the two allocation vectors is equal to zero, then the unobserved components of utility for the two origin-destination dyads are uncorrelated.

[18]The choice probability in (3) can also be expressed as $p_{ijk} = \sum_m p_{ijk|b_m} p_{ijb_m}$, where $p_{ijk|b_m}$ is the probability of opting for destination k conditional upon choosing a destination belonging to the nest b_m, and p_{ijb_m} is the probability of choosing a destination in the nest b_m (Wen and Koppelman, 2001).

[19]Formally, this implies that there is a nest b_h such that $\alpha_{jjh} = 1$, and $\alpha_{jlh} = 0$ for all $l \in D$.

2.2 Migration flows and Multilateral Resistance to Migration

Imagine that individual migration decisions are observed over a set T of periods; the log of the scale of migration flows to country k at time $t \in T$ over the size of the population which opts for the origin country j, y_{jkt}, can be derived from the RUM model by averaging (5) over the set of individuals i. The result is given by:

$$y_{jkt} = \boldsymbol{\beta}' \left(\frac{\boldsymbol{x}_{jkt}}{\tau} - \boldsymbol{x}_{jjt} \right) + r_{jkt} + \eta_{jkt} \tag{6}$$

The error term η_{jkt} is orthogonal to $\boldsymbol{x}_{jkt}$ and $\boldsymbol{x}_{jjt}$, serially uncorrelated, and independently and identically distributed over the set of origin-destination pairs, and r_{jkt} is equal to:

$$r_{jkt} = \ln \left(\sum_m (\alpha_{jkm})^{1/\tau} \left(\sum_{l \in b_m} (\alpha_{jlm} e^{V_{jlt}})^{1/\tau} \right)^{\tau-1} \right) \tag{7}$$

The term r_{jkt} in (7) represents the multilateral resistance to migration, as it captures the influence exerted by the opportunities to migrate to other destinations upon migration from country j to country k at time t. Taking the partial derivative of r_{jkt} with respect to the deterministic component of utility in destination l we get:

$$\frac{\partial r_{jkt}}{\partial V_{jlt}} = \sum_n \left(w_{jkn} \frac{\tau-1}{\tau} (p_{jlt}|b_n) \right) \leq 0 \tag{8}$$

where:

$$w_{jkn} = \frac{(\alpha_{jkn})^{1/\tau} \left(\sum_{g \in b_n} (\alpha_{jgn} e^{V_{jgt}})^{1/\tau} \right)^{\tau-1}}{e^{r_{jkt}}}$$

The multilateral resistance to migration term r_{jkt} is always a non-increasing function of V_{jlt}, and the inequality in (8) is equal to zero only if $\boldsymbol{\alpha}_{jk}' \boldsymbol{\alpha}_{jl} = 0$. An increase in V_{jlt} redirects towards l proportionally more individuals that would have opted for destination k than individuals who would have stayed in the country of origin j, thus reducing the bilateral migration rate y_{jkt} in (6).

3 Estimation strategy

The distribution of the stochastic term ϵ_{ijk} in (1), which depends upon the specific assumptions about the GEV generating function, are closely related to the shape of the multilateral resistance to migration term r_{jkt} in (6). This section analyzes which are the specifications about the GEV generating function in (2) which justify the alternative econometric approaches that have been adopted in the literature, and it then introduces the more general specification adopted in this paper, and the ensuing econometric strategy.

3.1 The traditional approach

As recalled in the introduction, the traditional estimation approach in the migration literature assumes that the bilateral migration rate can be expressed as a function of origin and characteristics only (Hanson, 2010). This approach, which has been adopted by Clark, Hatton, and Williamson (2007), Pedersen, Pytlikova, and Smith (2008), Lewer and den Berg (2008), Mayda (2010) and Grogger and Hanson (2011), uses all the variability in the data to identify the vector of coefficients $\boldsymbol{\beta}$.[20]

In terms of our RUM model, this requires that no multilateral resistance to migration term r_{jkt} appears in the equation to be estimated. Going back to (7), this happens if and only if the allocation matrix $\boldsymbol{\alpha}_j$ is a $n(j) \times n(j)$ identity matrix, so that any location is entirely allocated to a singleton, and the multilateral resistance to migration term r_{jkt} which appears in (6) is identically equal to zero.

This assumption on the allocation matrix implies that the underlying GEV generating function defined in (2) simplifies to:

$$G^1(Y_{ij1}, ..., Y_{ijn(j)}) = \sum_{l \in D_j} Y_{jl} \tag{9}$$

The function G^1 in (9) entails that ϵ_{ijk} in (1) follows an Extreme Value Type-1 distribution (McFadden, 1974), and it generates the choice probabilities that identify the multinomial logit model:

---

[20]When the dataset has a longitudinal dimension, the inclusion of origin dummies removes the variability across origins, but the identification of $\boldsymbol{\beta}$ still comes from the variability over time for each origin.

$$p_{ijk} = \frac{e^{V_{jk}}}{\sum_{l \in D_j} e^{V_{jl}}} \tag{10}$$

The multinomial logit model is characterized by the Independence of Irrelevant Alternatives,[21] as the relative probability of opting for two destinations is independent from the attractiveness, or even the existence, associated to any other destination: an increase in the attractiveness of another destination draws proportionally from all the other destinations, so that relative choice probabilities remain unchanged.[22]

Train (2003) observes that the distribution of the stochastic component ϵ_{ijk} "is not defined by the choice situation *per se*", and IIA can actually "be interpreted as a natural outcome of a well-specified model". Still, data constraints are often binding in the migration literature, and they can induce to opt for a parsimonious specification of the location-specific utility, so that it is relevant to explore identification strategies which can accommodate for a correlation in unobservables across alternatives, which in turn implies that the multilateral resistance to migration term r_{jkt} is present in the equation to be estimated.

3.2 The inclusion of origin-time dummies

While the traditional approach made full use of the variability across destinations and origins and over time in the data to identify the vector of coefficients $\boldsymbol{\beta}$, Ortega and Peri (2009) have reduced the amount of variability used for identification through the inclusion of origin-time dummies.

The identification strategy adopted in Ortega and Peri (2009) is consistent with their proposed underlying RUM model, which generalizes the one in Grogger and Hanson (2011) by "allowing for unobserved individual heterogeneity between migrants and non-migrants". The inclusion of origin-time dummies makes their estimation approach consistent with the discrete choice model which is produced by the following GEV generating function:

---

[21]The multinomial logit choice probabilities in (10) were originally derived by Luce (1959) from the IIA property, which represented a corollary of a set of axioms about the choice over discrete alternatives that he had proposed; Debreu (1960) provided an early critique of the plausibility of the IIA property.

[22]Grogger and Hanson (2011) verify that the estimated coefficient for the income differential remains stable when destinations are removed from the choice set of prospective migrants, as a violation of the IIA assumption would entail instability of the estimated coefficients (Hausman and McFadden, 1984).

$$G^2(Y_{ij1}, ..., Y_{ijn(j)}) = Y_{ij1} + \left(\sum_{l \in D} Y_{jl}^{1/\tau} \right)^{\tau} \tag{11}$$

which can be derived from (2) assuming that the allocation matrix $\boldsymbol{\alpha}_j$ is the following $n(j) \times 2$ matrix:

$$\boldsymbol{\alpha}_j{}' = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 1 & ... & 1 \end{pmatrix} \tag{12}$$

The allocation matrix in (12) implies that the two nests represent a partition of the set D_j, as all the destinations in D are entirely allocated to the same nest, while the origin j belongs to a singleton. The GEV generating function G^2 gives rise to the choice probability corresponding to the nested logit (McFadden, 1978), and it also implies that the multilateral resistance to migration term r_{jkt} in (7) can be simplified to:

$$r_{jkt} = (\tau - 1) \ln \left(\sum_{l \in D} e^{V_{jlt}/\tau} \right) \tag{13}$$

The key characteristic of (13) is that it is invariant across destination countries for a given time t. Hence, the inclusion of origin-time dummies suffices to control for multilateral resistance to migration when the discrete choice probabilities are generated by the function in (11). This reduces the variability that is used to identify $\boldsymbol{\beta}$, to remove the influence of the time-varying component of multilateral resistance to migration in (13).

When the dataset only has one either cross-sectional or longitudinal dimension, (13) also entails that the inclusion of either origin or time dummies suffices to make the identification strategy consistent with the specific violation of IIA induced by the GEV generating function G^2. This implies that the estimates provided in Beine, Docquier, and Ozden (2011), who assume that the stochastic components of their RUM model follows an Extreme Value Type-1 distribution, and in Theoharides, McKenzie, and Yang (2010) can be consistent even if IIA is violated in this specific way.

The inclusion of origin-time dummies among the controls implies that the underlying pattern of substitution across alternative locations is richer than in the traditional approach: an increase in the attractiveness of destination l can draw from another destination k more than it does from the origin country j, so that the bilateral migration rate y_{jkt} falls.[23]

---

[23]This approach shares a key feature with the traditional approach, as the sorting of migrants across

3.3 A more general approach

Let us go back to the general specification for the multilateral resistance to migration term r_{jkt}, which is produced by the more general GEV generating function G in (2), with no restrictions on the size and composition of the allocation matrix $\boldsymbol{\alpha}_j$.[24] We reproduce here the general expression (7) of r_{jkt}:

$$r_{jkt} = \ln \left(\sum_m (\alpha_{jkm})^{1/\tau} \left(\sum_{l \in b_m} (\alpha_{jlm} e^{V_{jlt}})^{1/\tau} \right)^{\tau-1} \right)$$

Differently from Ortega and Peri (2009), the term r_{jkt} varies across destinations, as these can be allocated unevenly across different nests. Hence, the inclusion of origin-time dummies would not suffice to control for multilateral resistance to migration. Consider also that r_{jkt} is unobservable for the econometrician, as it depends (i) on the value of deterministic component of location-specific utility for countries other than j and k, and (ii) on the unobserved allocation matrix $\boldsymbol{\alpha}_j$, which reflects unknown preferences of prospective migrants.

The equation to be estimated is then:

$$y_{jkt} = \boldsymbol{\beta}' \left(\frac{\boldsymbol{x}_{jkt}}{\tau} - \boldsymbol{x}_{jjt} \right) + \varepsilon_{jkt}$$

where:

$$\varepsilon_{jkt} = r_{jkt} + \eta_{jkt} \tag{14}$$

The multilateral resistance to migration r_{jkt} entails that the error term ε_{jkt} in (14) is not well-behaved. Specifically, r_{jkt} will be, in general, serially correlated, as the resistance to migration exerted by other destinations is likely to evolve slowly over time, and spatially correlated across origin-destination dyads.

With respect to spatial correlation, observe that r_{jkt} will be in general correlated with r_{jlt}: the bilateral migration rates from the same origin country j to the two destinations k and l

---

destinations l and k is still insensitive to a variation in the attractiveness of a third destination $g \in D$.

[24] As observed by McFadden (2001), "tractable versions [of GEV models] fall short of being able to represent all RUM-consistent behavior", but the discrete choice model produced by the specific GEV generating function introduced by Wen and Koppelman (2001) and used in this paper represents the least restrictive used so far in the migration literature.

will be both influenced by the attractiveness of migration to other alternative destinations. By the same token, in general we will also have that r_{jkt} will be correlated with r_{hkt}: the bilateral migration rates from two different origins j and h to the same destination k will both be affected by the attractiveness of migration to other alternative destinations. Multilateral resistance to migration induces spatial correlation not only for the flows towards various destinations from the same origin country, but also for the flows originating from different origins and directed to the same destination country.[25]

When the error term is serially and spatially correlated, OLS still provides consistent estimates of the coefficients $\boldsymbol{\beta}$ (Driscoll and Kraay, 1998), but the standard errors will be incorrect. Driscoll and Kraay (1998) propose an approach to estimate the standard errors of the coefficients which is robust to non-spherical errors, and that be implemented following Hoechle (2007).

Still, the approach by Driscoll and Kraay (1998) addresses only some of the challenges posed by multilateral resistance to migration, as it requires exogeneity of the regressors. But the presence of r_{jkt} in the error term is likely to violate the exogeneity assumption, as r_{jkt} can be correlated with the regressors.

To get an intuition of the endogeneity problem due to multilateral resistance to migration, consider a likely key macro determinant of the scale of migration flows, namely GDP per capita at origin, which enters the vector $\boldsymbol{x}_{jjt}$. GDP per capita at origin j can correlate with GDP per capita in some of the destination countries, which are included in r_{jkt}; this can occur because of the exposure to common economic shocks, or because of a partial business cycle synchronization due to trade and investment flows.

We can also consider the case where visa policy at destination enters the vector $\boldsymbol{x}_{jkt}$. Visa policies - which can exert a substantial influence on the scale of bilateral migration flows (Bertoli, Fernández-Huertas Moraga, and Ortega, 2011) - can be coordinated at the supranational level. For instance, the list of third countries whose nationals need a visa to enter the European Union is determined by the European Council: when a country is included in this list, a simultaneous change in the bilateral visa policies towards this country adopted by EU member states is observed. As far as EU countries are perceived as close substitutes by would-be migrants from third countries, we have that $\boldsymbol{x}_{jkt}$ correlates with r_{jkt}.

---

[25]This, in turn, implies that multilateral resistance to migration can represent a challenge for the econometric analysis even if, as in Clark, Hatton, and Williamson (2007), the data relate to flows to a single destination.

These arguments entail that we need an estimator that is also able to handle the endogeneity of the regressors.[26]

3.3.1 The multifactor error structure in Pesaran (2006)

Pesaran (2006) deals with the challenges connected to the estimation of the following panel data model:

$$y_{it} = \boldsymbol{\delta}_i' \boldsymbol{d}_t + \boldsymbol{\beta}' \boldsymbol{x}_{it} + \epsilon_{it} \tag{15}$$

where:

$$\epsilon_{it} = \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \eta_{it} \tag{16}$$

The error term has a multifactor structure,[27] as it contains the inner product between a vector $\boldsymbol{\gamma}_i$ of panel-specific factor loadings, and a vector $\boldsymbol{f}_t$ of time-varying factors. Pesaran (2006) allows the error term ϵ_{it} to be heteroskedastic[28], serially and spatially correlated, and correlated with the regressors, and it proposes a consistent estimator for the coefficient vector $\boldsymbol{\beta}$ which does *not* require to know the dimension of the vector $\boldsymbol{f}_t$, nor the elements in the vector $\boldsymbol{\gamma}_i$.

Here, we want to show that the multilateral resistance to migration term r_{jkt}, which enters the equation to be estimated, can be approximated in a way that fits the multifactor error structure in (16). Let $\widetilde{V}_{jl}$ the dyad-specific average of the deterministic component of utility V. Relying on a Taylor expansion around $\widetilde{V}_{jl}$, we can approximate the multilateral resistance to migration term r_{jkt} introduced in (7) as follows:

$$r_{jkt} \approx \widetilde{r}_{jk} + \sum_n \sum_{l \in D} \left[\left(w_{jkn} \frac{\tau - 1}{\tau} \right) \left((\widetilde{p}_{jlt}|b_n)(V_{jlt} - \widetilde{V}_{jl}) \right) \right] \tag{17}$$

where $\widetilde{r}_{jk}$ and $\widetilde{p}_{jlt}|b_n$ represent respectively the value of r_{jkt} and of the probability $p_{jlt}|b_n$ of opting for destination l conditional upon the choice of the nest b_n in correspondence to the dyad-specific averages $\widetilde{V}_{jl}$. Observe that the first term of the product within square brackets

---

[26]The use of external instruments is hardly an option here, as endogeneity is not confined to a regressor, but to *all* relevant determinants of the scale of migration flows.

[27]Bai (2009) refers to the same structure of the error term as the interactive fixed effects model.

[28]Even if we do not derive our estimated equation from a log-linearization, this allows us to fully address the challenges posed by heteroskedasticity which are detailed by Santos Silva and Tenreyro (2006).

in (17) is dyad-specific, while the second term is time-specific; hence, using vector notation, we can rewrite (17) more compactly as:

$$r_{jkt} \approx \widetilde{r}_{jk} + \boldsymbol{\gamma}_{jk}{}' \boldsymbol{f}_t \tag{18}$$

The elements in the vector of dyad-specific factor loadings $\boldsymbol{\gamma}_{jk}$ depend on the unobservable preferences of individuals from origin j, which are reflected in the allocation matrix $\boldsymbol{\alpha}_j$, as well as upon the unknown dissimilarity parameter τ, while the elements in the vector $\boldsymbol{f}_t$ are an affine function of the deterministic component of location-specific utility.

Using (18), we can rewrite the equation to be estimated as:

$$y_{jkt} = \boldsymbol{\beta}_1{}' \boldsymbol{x}_{jkt} + \boldsymbol{\beta}_2{}' \boldsymbol{x}_{jjt} + \beta_{jk} d_{jk} + \epsilon_{jkt} \tag{19}$$

where d_{jk} is a dummy for the dyad (j,k), and $\epsilon_{jkt} = \boldsymbol{\gamma}_{jk}{}' \boldsymbol{f}_t + \eta_{jkt}$, and the vectors of coefficient to be estimated are related to the parameters in the RUM model as follows: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}/\tau$ and $\boldsymbol{\beta}_2 = -\boldsymbol{\beta}$.

3.3.2 The Common Correlated Effects estimator

The presence of a multifactor error structure which correlates with the regressors implies that OLS or FE estimates of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in (19) will be inconsistent. Pesaran (2006) proposes an alternative estimator: the Common Correlated Effects (CCE) estimator, which is able to control for the unobserved multifactor component of the error term. In terms of the equation derived from our underlying RUM model, the CCE estimator allows us to recover a consistent estimate of the effects of the determinants of bilateral migration flows without having to assume that IIA holds, and allowing for a more general violation of IIA than the one considered in Ortega and Peri (2009).

Pesaran (2006) demonstrates that $\boldsymbol{\gamma}_i{}' \boldsymbol{f}_t$ in (15) can be expressed as a dyad-specific linear combination of the cross-sectional averages of the dependent variable and of the regressors. Specifically, he demonstrates that a consistent estimate of $\boldsymbol{\beta}$, $\boldsymbol{b}_{CCE}$, can be obtained from the estimation, through OLS, of the following regression:

$$y_{jkt} = \boldsymbol{\beta}_1{}' \boldsymbol{x}_{jkt} + \boldsymbol{\beta}_2{}' \boldsymbol{x}_{jjt} + \beta_{jk} d_{jk} + \boldsymbol{\lambda}_{jk}{}' \widetilde{\boldsymbol{z}}_t + \eta_{jkt} \tag{20}$$

where the vector of auxiliary regressors $\widetilde{\boldsymbol{z}}_t$ is equal to:

$$\widetilde{\boldsymbol{z}}_t = \frac{1}{\sum_{(j,k)} \omega_{jkt}} \Big(\sum_{(j,k)} \omega_{jkt} y_{jkt}, \sum_{(j,k)} \omega_{jkt} \boldsymbol{x}_{jkt}{}', \sum_{(j,k)} \omega_{jkt} \boldsymbol{x}_{jjt}{}' \Big)'$$

and ω_{jkt} is the weight assigned to each origin-destination dyad at time t in the estimation. The consistency of $\boldsymbol{b}_{CCE}$ is established by Pesaran (2006) by demonstrating that $\boldsymbol{\lambda}_{jk}{}'\widetilde{\boldsymbol{z}}_t$ converges in quadratic mean to $\boldsymbol{\gamma}_{jk}{}'\boldsymbol{f}_t$ as the cross-sectional dimension of the panel goes to infinity, with the longitudinal dimension being either fixed or also diverging to infinity (Pesaran, 2006). Monte Carlo simulations in Pesaran (2006) also show the good finite sample properties of the CCE estimator, which already produces satisfactory results when $N = 30$ and $T = 20$. Pesaran and Tosetti (2011) confirm these properties even when η_{jkt} is serially or spatially correlated.

3.3.3 Multilateral resistance to migration and the CCE estimator

Some key features of the CCE estimator proposed by Pesaran (2006) are worth emphasizing in relationship with its application to the estimation of the determinants of bilateral migration rates.

First, it does not require to know the dimension of the vector of time-specific common shocks which enters the error term. This fits nicely with our general RUM model, as different specifications of the allocation matrix $\boldsymbol{\alpha}_j$ translate into a different size of the vector $\boldsymbol{f}_t$ which approximates the multilateral resistance to migration term r_{jkt}. This allows us to obtain estimates of the vector of coefficients $\boldsymbol{\beta}$ without having to introduce additional assumptions on $\boldsymbol{\alpha}_j$.

Second, the CCE estimator allows us to identify the effects of determinants of bilateral migration rates which are specific to each origin country, such as GDP per capita. This further differentiates our approach from Ortega and Peri (2009), as the inclusion of origin-time dummies, which is not consistent with a more general GEV generating function, prevents the identification of the effects of relevant push factors of migration flows.

Third, we do *not* need to have data on multiple destinations to be able to control for multilateral resistance to migration with the CCE estimator. Recall, from (17) and (18), that the r_{jkt} term is an affine function of the deterministic component of utility V_{jlt} for the same origin country j. So, a legitimate question arises: is it possible to control for multilateral resistance to migration even when the data refer to a cross-section of origins, but to a single destination? The answer to this question is positive, and it relates to the discussion about

the pattern of spatial correlation induced by multilateral resistance to migration discussed in Section 3.3.

The pattern of correlation in the error term, not only across destinations but also across origins, contains information about the unobserved attractiveness of other destinations, and to the related unobserved bilateral flows. Intuitively, once one controls for the observed determinants of bilateral flows, residual simultaneous variations in the flows to a given destination from the origin countries included in the sample are acting as a *mirror*, reflecting the effects of changes in the opportunities to migrate to other unobserved destinations. The efficacy of such a mirror effects depends on the similarity of the structure of preferences across different origins, as reflected in the allocation matrix $\boldsymbol{\alpha}_j$, and on the correlation between the attractiveness of various destinations. Similarity in the preferences across origin countries or correlation in the deterministic component of utility across destinations imply that the cross-sectional averages of the dependent and of the independent variables referring *only* to other origin countries which enter the vector of auxiliary regressors $\widetilde{\boldsymbol{z}}_t$ provides us with the information that is needed to control for the influence exerted by multilateral resistance to migration on bilateral flows.

4   Data and descriptive statistics

Our dataset has three main components: migration flows to Spain in the 1997-2009 period; migration policies in Spain during the same period; and quarterly real GDP series for the countries of origin of migrants to Spain. Here, we first present each of these components and we look at their main characteristics, then we provide the relevant descriptive statistics.

4.1   Migration flows

The migration flows data come from the *Estadística de Variaciones Residenciales* (EVR). This is an administrative dataset collected by the Spanish *Instituto Nacional de Estadística* (INE). The EVR gathers all the variations in the municipal registry (*Padrón Municipal de Habitantes*) throughout the year: each observation in the EVR corresponds either to an inscription in or to a cancelation from the *Padrón*, and it includes information on the date in which the variation occurred, and on the age, gender and country of birth of the individual to whom the variation refers. We use the observations referring to the first inscription of

Figure 1: Monthly Immigration Inflows to Spain 1997-2009 (EVR)
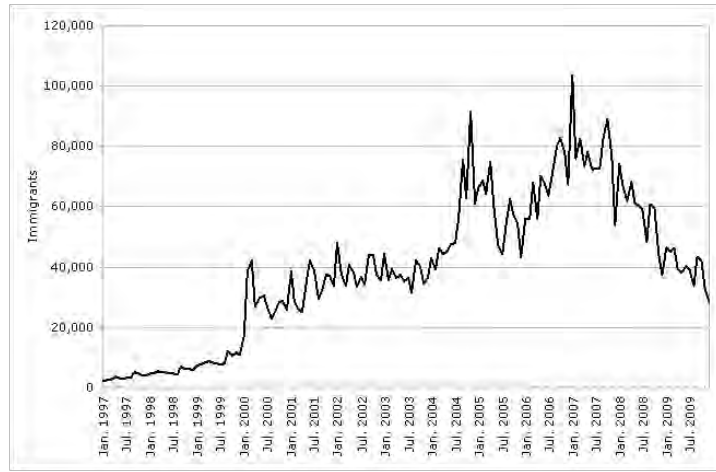
foreign-born individuals coming from abroad in the *Padrón* to measure immigration flows to Spain: the EVR contains 6,166,133 of these observations between January 1997 and December 2009,[29] related to individuals from 208 countries of origin.[30]

By restricting our attention to inscriptions of foreign-born individuals coming to Spain from abroad, we are obtaining an almost perfect measure of gross immigration inflows. The measure would be perfect if every individual registered immediately upon arrival. Although registration is not mandatory, most immigrants eventually do register, independently of their legal status, as registration gives them access to all basic municipality services, most notably free health care and education (Bertoli, Fernández-Huertas Moraga, and Ortega, 2011). The Appendix A.1 discusses in detail the accuracy of the EVR in measuring immigration flows to Spain, comparing EVR figures with those that can be obtained from alternative data sources.

Figures 1 and 2 plot the monthly and quarterly series of immigration flows to Spain over our period of analysis according to the EVR. Despite the large apparent variability in the overall immigration series, there does not seem to be relevant seasonal patterns in the data. None is found if we regress quarterly data on year and quarter dummies: the quarterly dummies are not significant. For the monthly data, a regression on year and month dummies shows the months of August and December as those in which registrations are significantly

---

[29]As recalled in the introduction, these figures correspond to an unprecedented - even from an international perspective (OECD, 2010) - surge in immigration.

[30]The EVR also codifies some former states, such as the USSR or Yugoslavia.

19

Figure 2: Quarterly Immigration Inflows to Spain 1997-2009 (EVR)
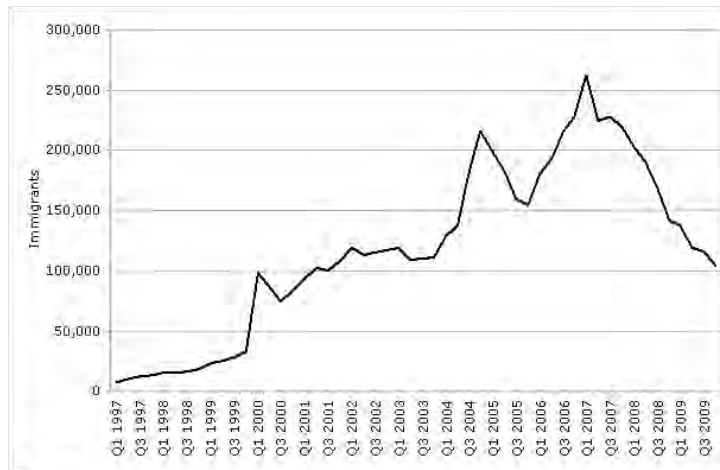
Figure 3: Total flows and excluding immigrants from Bulgaria and Romania (1997-2009)
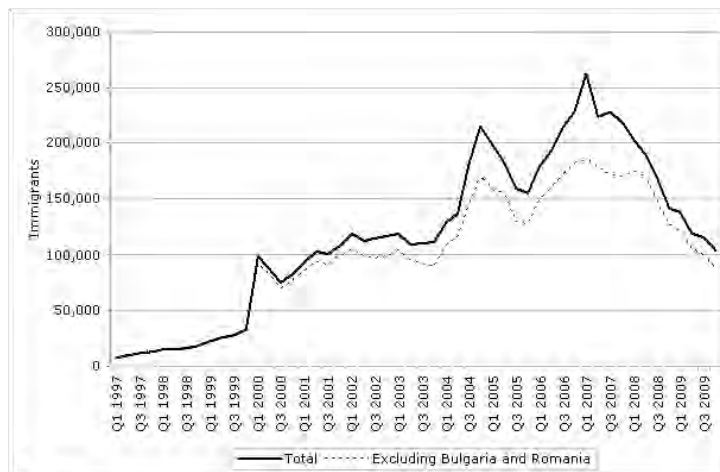
lower (between 15 and 20 percent) than in the rest of the year, coinciding with the summer and winter holidays in Spain.

There are three noticeable spikes in the series: the first one corresponds to the January 2000 law that ensured access to basic services for those registered (*Ley Orgánica* 4/2000); the second one can be associated to the 2005 massive amnesty and happened in November 2004; finally, the third one has to do with the accession of Romania and Bulgaria to the EU in January 2007, taking into account that Romanians have created the largest immigrant community in Spain (see Figure 3 for the evolution of total flows excluding the two most recent EU member states).

Our analysis aggregates the EVR data at the quarterly level, as this is the finest period of time for which we can gather information on the economic conditions at origin. We restrict our sample to the origin countries with a positive total number of immigrants in all the 52 quarters included in our period of analysis: 98.6 percent of total migration flows to Spain between January 1997 and December 2009 originated from these countries,[31] whose population represents 86 percent of the world total.

In our empirical analysis below, our dependent variable will be the log of the emigration rate to Spain from a given origin country over a quarter, consistently with the model presented in Section 2. This is calculated as the total number of immigrants to Spain from origin country j who registered during a given quarter divided by the population of that country of origin j in that year.[32]

4.2   Spanish migration policies

We gather data on Spanish migration policies between 1997 and 2009; specifically, we codify the following policies which are likely to influence bilateral migration flows in the EVR: (i) general policies - the 2000 Amnesty, the 2005 Amnesty; (ii) bilateral policies - visa agreements, double nationality agreements, social security agreements, agreements on the signature of labor contracts at origin; and (iii) multilateral treaties - membership to the EU-15, membership to the Schengen area, 2004 EU enlargement, 2007 EU enlargement. The Appendix A.2 describes the definition and sources of these variables.

Our database comprises 8 EU-wide agreements transposed into Spanish Law through Decrees,[33] 48 national Laws, Resolutions and Orders dealing with migration issues,[34] and 94 bilateral agreements between Spain and origin countries regarding matters such as the need of a visa to enter Spain, portability of social security benefits and the legal recognition of

---

[31] The share of the observations where the recorded migration flow is equal to zero is much lower than in the dataset employed by Beine, Docquier, and Ozden (2011), where it stands at 36 percent; Beine, Docquier, and Ozden (2011) assess the sensitivity of their estimates to the inclusion of these zero observations, and they validate the estimates obtained from the specifications where these observations are dropped from the sample as "results are highly robust to various econometric techniques accounting for the large proportion of zeros". A similar conclusion is reached also by Grogger and Hanson (2011).

[32] Our population figures are taken from the World Development Indicators (World Bank, 2010), and vary only at the yearly level.

[33] The EU enlargement to 25 members that applied from May 1, 2004 is one such entry in our database

[34] These include, for example, the 2005 amnesty that applied from February 7, 2005 to May 7, 2005.

educational degrees. We have taken the data from the web pages of the Ministry for Labor and Immigration and the *Boletín Oficial del Estado*, a daily official bulletin where all Spanish legislation is published.

We model these migration policies as dummy variables that change from 0 to 1 from the month the policy is applied. For instance, the 2000 Amnesty is modeled as a 0 before January 2000 and as a 1 afterwards. Another example, already studied by Bertoli, Fernández-Huertas Moraga, and Ortega (2011) is the bilateral agreement between Ecuador and Spain regarding the need of a visa for Ecuadorians to enter Spain. We model this as a dummy taking value 1 when a visa is needed to enter Spain and value 0 otherwise; in the Ecuadorian case, this means the value of the visa dummy is 0 before August 2003 and 1 after that date. We present a more detailed description of the construction of the dataset in the Appendix A.2.

This set of ten variables is able to explain, in a simple OLS regression, up to 54 percent of the total variation on the log of the monthly or quarterly emigration rates to Spain by country of origin. This shows that our migration policy specification has a good deal of variability and potential explanatory power.

4.3 Economic conditions at origin

Our estimation strategy requires the use of high-frequency data, and we were able to gather quarterly real GDP data for 61 origin countries, representing 87 percent of total migration flows to Spain over the 1997-2009 period. As detailed in the Appendix A.3, our data sources are the International Financial Statistics (IMF, 2010a), the April 2010 issue of the World Economic Outlook (IMF, 2010b) and the data published by some Central Banks.

We divide our quarterly real GDP series by the yearly population figures from the World Development Indicators (World Bank, 2010) to obtain real GDP per capita series that we use as a proxy for the time-varying economic conditions at origin. Since the series vary widely in terms of base year, adjustments on seasonality, base currency and other aspects, we construct a country-specific seasonally-adjusted real GDP per capita index (setting the index equal to 100 in the first quarter of 2000). The raw correlation between the log of the GDP per capita index by quarter and country of origin and the log of the emigration rate to Spain is 0.05. In a simple regression of the two variables, the coefficient on the GDP per capita index is 0.8 and is only able to explain 0.3 percent of the variation in quarterly

Figure 4: Quarterly Immigration Inflows to Spain, total and selected sample (1997-2009)
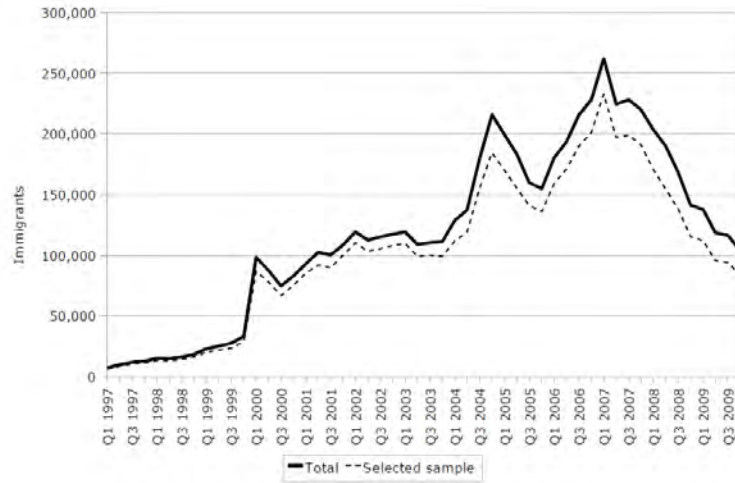
emigration rates.

4.4   Summary statistics

When combining our migration flows, migration policies and real GDP per capita datasets, we are left with 3,020 observations. Out of the 6,166,133 immigrants who, according to the EVR, entered Spain between January 1997 and December 2009 coming from 208 countries, we keep in our sample 5,341,586 immigrants coming from 61 countries, which host 51 percent of the world population. Figure 4 shows that these 61 countries keep the basic time series structure of the overall number of immigrants.

We present in Table 1 some summary statistics of this emigration rate (expressed in migrants to Spain per 1,000,000 inhabitants) and of the GDP per capita index in our sample. In order to allow a straightforward comparison, we also construct a country-specific index for emigration rates. We weight observations by the population of the country of origin since we are interested in exploring determinants of emigration rates over the whole population.

Table 1 shows that the variability is much more substantial in the emigration rate than in the GDP per capita during the period. The mean emigration rate per quarter to Spain was 32.88 emigrants per 1,000,000 inhabitants with a maximum in the sample of 3,099 emigrants in the first quarter of 2007 from Romania and a minimum of 0.01 in the first quarter of 1997 from Indonesia. For the country-specific index, the average of 268 reflects the growth in migration rates from 2000. The relative maximum (15,740) corresponds to Paraguay in

23

Table 1: Summary statistics

Variable	mean	s.d.	min	max	obs.
Emigrants to Spain per 1,000,000 inhabitants	32.88	136.75	0.01	3,098.78	3,020
Emigration rate index (2000q1=100)	267.58	381.83	0.30	15,470.04	3,020
Real GDP per capita index (2000q1=100)	115.17	19.91	69.61	223.34	3,020
January 2000 Amnesty	0.83	0.37	0	1	3,020
November 2004 Amnesty	0.44	0.49	0	1	3,020
EU-15	0.11	0.31	0	1	3,020
Schengen Area	0.09	0.28	0	1	3,020
EU May 2004 Eastern Enlargement	0.01	0.10	0	1	3,020
EU May 2007 Romania and Bulgaria Enlargement	0.002	0.05	0	1	3,020
Visa requirement	0.57	0.50	0	1	3,020
Bilateral Agreement on Nationality	0.05	0.23	0	1	3,020
Bilateral Agreement on Social Security	0.13	0.33	0	1	3,020
Bilateral Agreement on Contracts at Origin	0.02	0.13	0	1	3,020

Note: quarterly series on 61 countries (1997-2009), all descriptive statistics are weighted by population at origin; see the Appendix A.2 for a description of the immigration policy variables.

the first quarter of 2007 whereas the minimum (0.30) is Ecuador in the first quarter of 1997. For the GDP per capita index, the average value (weighted by population) in the sample is 115 with a minimum of 70 for Venezuela in the first quarter of 2003 and a maximum of 223 for Georgia in the second quarter of 2008. We can observe the scatter-plot of the log of both indexes in Figure 5.

It is perhaps more informative to look directly at the time series evolution of the variables the way they will be used in the empirical analysis below. The following series of figures present this representation for the log of the emigration rate and the log of real GDP per capita for the four top emigrant sending countries to Spain during the period: Romania (809,857 emigrants), Morocco (666,798 emigrants), Ecuador (490,580 emigrants) and Colombia (377,780 emigrants).

Figures 6 to 9 show that, despite a general upward time trend in most of the series that the empirical analysis will have to account for, there is substantial time and cross-sectional variation to be exploited in the dataset.

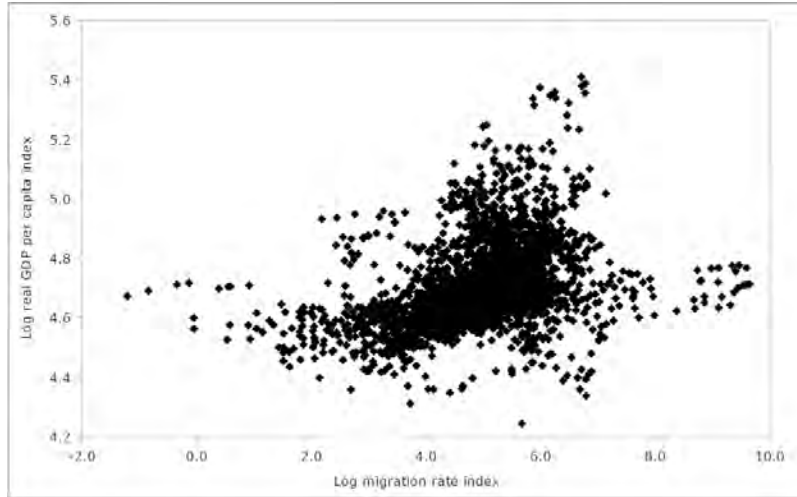Figure 5: Emigration and GDP at origin, selected sample (1997-2009)
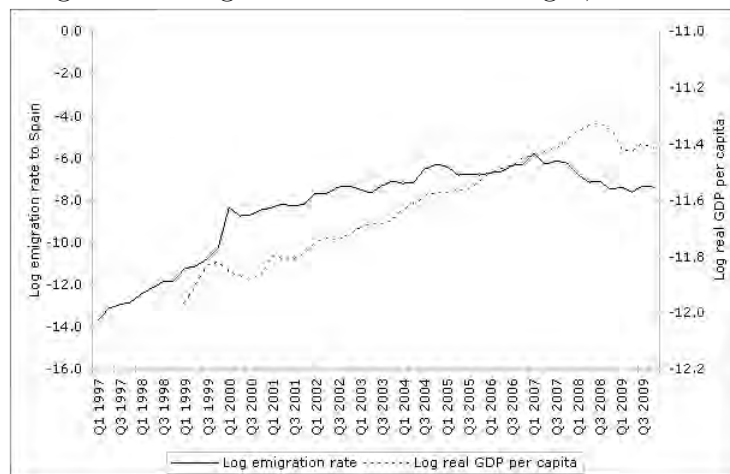

Figure 6: Emigration and GDP at origin, Romania

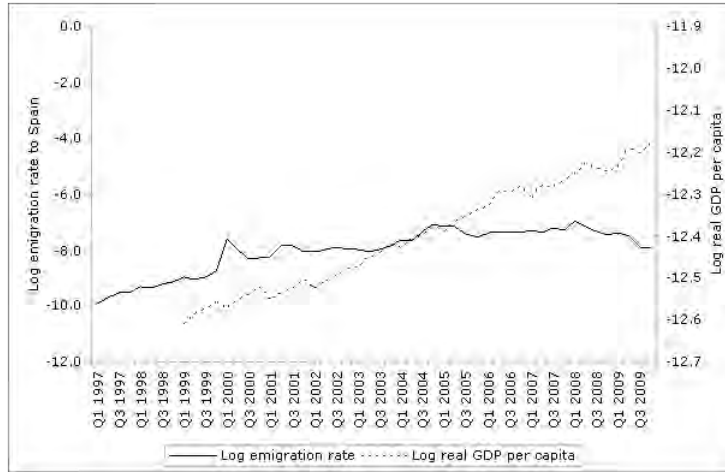Figure 7: Emigration and GDP at origin, Morocco

Figure 8: Emigration and GDP at origin, Ecuador
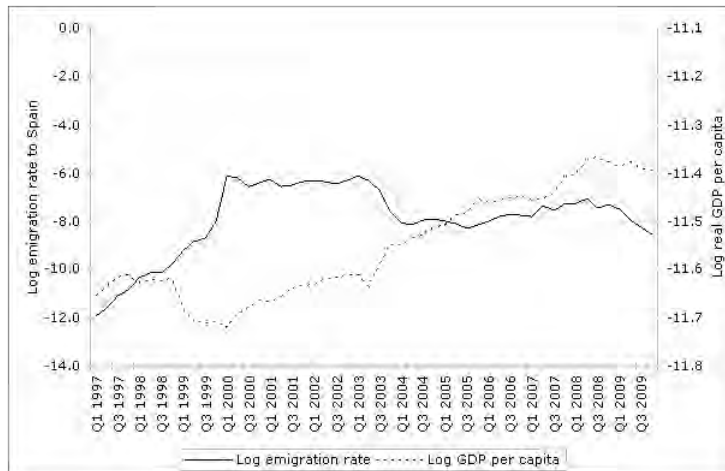
Figure 9: Emigration and GDP at origin, Colombia
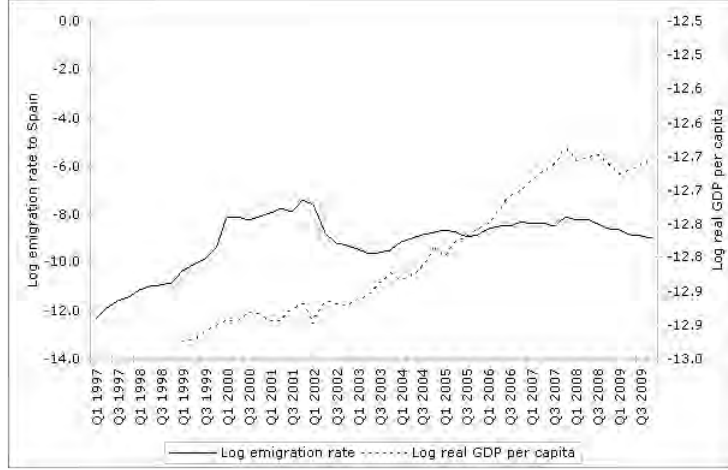
5 Econometric analysis

The econometric analysis of the determinants of bilateral migration flows to Spain over 1997-2009 follows the steps entailed by the estimation strategy outlined in Section 3. We report here the equation to be estimated, derived on the basis of the RUM model presented in Section 2:

$$y_{jkt} = \boldsymbol{\beta}_1{}' \boldsymbol{x}_{jkt} + \boldsymbol{\beta}_2{}' \boldsymbol{x}_{jjt} + \beta_{jk} d_{jk} + \boldsymbol{\lambda}_{jk}{}' \widetilde{\boldsymbol{z}}_t + \eta_{jkt}$$

Consistently with the model, the dependent variable y_{jkt} is represented by the log of the quarterly migration rate to Spain for each of the 61 origin countries included in our sample. The vector $\boldsymbol{x}_{jkt}$ contains a number of dyad-specific elements, represented by the bilateral immigration policies and multilateral treaties described in Section 4.2, while we control for all origin-invariant factors - such as the level of GDP or unemployment at destination - through the inclusion of quarter fixed effects, and for all time-invariant factors - such as cultural or linguistic proximity - through origin fixed effects. The vector $\boldsymbol{x}_{jjt}$ includes (various lags of) the log of real GDP per capita at origin, and origin-year fixed effects to control for all unobserved origin- and dyad-specific time-varying determinants of bilateral migration flows.[35]

With respect to our measure of GDP at origin, we include lagged values given that we

---

[35]The origin-year fixed effects also render our GDP per capita and emigration rate series stationary although the CCE estimator can accommodate unit roots.

have high-frequency migration data, and one can reasonably assume that would-be migrants do not react instantaneously to changes in economic conditions. We relied on the Akaike and Bayesian Information Criteria, and on Likelihood Ratio tests in order to select the optimal lag structure for each specification as suggested in Canova (2007), thus avoiding *ad hoc* choices. The optimal number of lags selected was four with all methods.

As a first step, we assume, as in Grogger and Hanson (2011) and Beine, Docquier, and Ozden (2011) that the stochastic term in the individual location-specific utility follows an Extreme Value Type-1 distribution, so that multilateral resistance to migration disappears, and (20) simplifies to:[36]

$$y_{jkt} = \boldsymbol{\beta_1}' \boldsymbol{x}_{jkt} + \boldsymbol{\beta_2}' \boldsymbol{x}_{jjt} + \beta_{jk} d_{jk} + \eta_{jkt}$$

This equation is estimated with a two-way error component model, and the results are presented in the first data column in Table 2. The model controls for origin-year fixed effects. The inclusion of this very rich structure of fixed effects allows us to control for those determinants of migration, such as demographic factors (Hanson and McIntosh, 2010a,b) or migrant networks (Munshi, 2003; Edin, Fredriksson, and Åslund, 2003; McKenzie and Rapoport, 2010; Beine, Docquier, and Ozden, 2011; Bertoli, 2010), which evolve at a pace that is slower than the frequency of our panel data. This substantially reduces the variability in the data that we are exploiting to identify the coefficient vector $\boldsymbol{\beta}$ but we are still able to precisely estimate the effect of GDP variations on migration decisions.

According to the first data column in Table 2, a 1.0 percent increase in real GDP per capita leads, after four quarters, to a 4.7 percent reduction in the migration rate to Spain.[37]

The estimates from this specification are consistent as long as multilateral resistance to migration does not influence bilateral migration flows to Spain. From Section 3, we know

---

[36]Observe that, as multilateral resistance to migration r_{jkt} does not enter the equation to be estimated, endogeneity should not be a pressing concern here: some of the crucial facets of the Spanish policy stance towards immigration are determined at the EU level and *bilateral* migration flows to Spain can be expected to exert only a very limited - if any - impact on economic conditions at origin. Remember that the largest emigration rate in our sample is 0.3 percent of the Romanian population in the first quarter of 2007. The median emigration rate in the sample is just 0.0002 percent.

[37]Note that this effect is notably larger than that found by Clark, Hatton, and Williamson (2007) for US immigration, which stands at 0.44; differently from them, we consider both legal and illegal immigration and exploit within-year variability in GDP. Our country-year fixed effects allow us to control for a much wider set of possible confounding factors that evolve slowly over time.

Dependent variable: log of quarterly emigration rate

Specification		(1)	(2)	(3)
Estimation methods		FE	FE	CCE
Regressors	*Lags*			
Log real GDP per capita	1	-1.57	-1.57	-1.57
		[0.24]***	[0.52]***	[0.29]***
	2	-1.05	-1.05	-0.46
		[0.26]***	[0.40]**	[0.29]
	3	-0.93	-0.93	-0.60
		[0.27]***	[0.52]*	[0.28]**
	4	-1.18	-1.18	-0.52
		[0.26]***	[0.62]*	[0.31]*
Visa requirement	0	-0.15	-0.15	-1.34
		[0.13]	[0.23]	[0.30]***
Other migration policy controls		yes	yes	yes
Quarter fixed effects		yes	yes	yes
Origin-year fixed effects		yes	yes	yes
Observations		2,776	2,776	2,776
Countries of origin		61	61	61
Frees' test (p-value)		2.71 (0.00)	-	-
Wooldridge's test (p-value)		32.64 (0.00)	-	-
Cross-sectional averages (p-value)		-	-	2.11 (0.00)
GDP per capita, cumulated effect	1	-1.57	-1.57	-1.57
		[0.24]***	[0.52]***	[0.29]***
	2	-2.63	-2.63	-2.02
		[0.28]***	[0.51]***	[0.35]***
	3	-3.56	-3.56	-2.62
		[0.33]***	[0.74]***	[0.43]***
	4	-4.74	-4.74	-3.14
		[0.37]***	[1.03]***	[0.53]***

Notes: standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1; observations
are weighted by population at origin; the number of lags of log real GDP per capita
has been determined according to AIC, BIC and LR tests to identify the optimal
lag structure following Canova (2007); specification (2) includes standard errors
computed following Driscoll and Kraay (1998); for specification (3), we present an
F-test that the coefficients on cross-sectional averages in the CCE estimator are
jointly zero, calculated on $F(659,1,378)$.

that this would induce spatial and serial correlation in the error term, and we follow Frees (1995) and Wooldridge (2002) to test for the presence of cross-sectional dependence and an autoregressive structure in the residuals.[38]

Table 2 shows that the null hypotheses of both tests are strongly rejected,[39] and this suggests that bilateral migration flows to Spain could be influenced by multilateral resistance to migration. This entails that the standard errors provide an incorrect basis for inference, and we re-estimated the same specification resorting to the method proposed by Driscoll and Kraay (1998) to obtain standard errors which are robust to serial and cross-sectional dependence in the error term.[40] The estimates in the second data column in Table 2 show that income at origin remains a significant determinant of bilateral migration flows - though the correction by Driscoll and Kraay (1998) substantially inflates its standard error, while the effect of the visa policy is still not significant. In Grogger and Hanson (2011), the effect of the visa waiver was marginally significant.

Still, these estimates are biased and inconsistent as multilateral resistance to migration is likely to make the regressors endogenous, as discussed in Section 3. Before resorting to the CCE estimator proposed by Pesaran (2006), which we have shown to be well-suited to address this specific form of endogeneity, it is interesting to consider the expected direction of the bias induced by multilateral resistance to migration with respect to the estimated coefficients of the GDP at origin and of the visa policy.

If real GDP per capita at origin correlates positively with real GDP per capita in some destinations that would-be migrants perceive as close substitutes to Spain, then the coefficient estimated in specifications (1)-(2) in Table 2 is *downward* biased. This occurs because an increase in GDP at origin is associated with an improvement in the attractiveness of other alternative destinations: if this is not controlled for, then the estimated effect of GDP at origin also captures the reduction in migration flows to Spain due to the increased attractiveness of other destination countries. This might be the relevant case with our dataset: Bertoli, Fernández-Huertas Moraga, and Ortega (2010) provided evidence that prospective

---

[38]We opted for the test for cross-sectional dependence proposed by Frees (1995) over the alternative test proposed by Pesaran (2004) as the latter could lack power and "miss out cases of cross-sectional dependence where the sign of the correlations is alternating" (De Hoyos and Sarafidis, 2006), because the multilateral resistance to migration term does not need to be positively correlated across different countries of origin.

[39]The two tests are implemented following De Hoyos and Sarafidis (2006) and Drukker (2003) respectively.

[40]The method by Driscoll and Kraay (1998) is implemented following Hoechle (2007).

migrants from the third largest origin country, Ecuador, regard Spain and the US as close substitutes, and the correlation between real GDP per capita in Ecuador and in the US stands at 0.54 once origin-year fixed effects are controlled for.[41] Of course, the direction of the bias depends on the prevailing pattern of correlation between the regressors and the omitted variables, so that the expectation is dependent on the characteristics of the data in the sample. If real GDP per capita at origin were to be negatively correlated with GDP per capita at alternative destinations, the bias would go in the opposite direction.

A similar line of reasoning suggests that the coefficient of the visa requirement estimated in specifications (1)-(2) in Table 2 is *upward* biased. A change in the Spanish visa policy towards one origin country occurs when also the other EU member states are adopting an identical change, when this decision follows a regulation by the European Council. An instance of such a change occurred in March 2001, when the citizens of the countries which were candidate to accession at that time were granted visa-free access to the EU by the EC Regulation No. 539/2001.[42] This regulation simultaneously changed the opportunities to migrate to other EU destinations: if this effect is not controlled for, then the estimated effect of the Spanish visa requirement also captures the increase in migration flows to Spain due to changes in the visa policy in other member states, biasing the negative coefficient upwards.[43]

The third data column in Table 2 presents the estimates obtained from the Common Correlated Effects estimator proposed by Pesaran (2006). As shown in the bottom panel of Table 2, the cross-sectional averages of the dependent and independent variables which are introduced as auxiliary regressors are jointly significant, which is required for the estimator to be valid.[44]

---

[41]The correlation between real per capita GDP in Colombia, the fourth largest origin country, and the US stands at 0.73, strengthening the expectation about the direction of the bias from neglecting multilateral resistance to migration. The correlations between other top destinations and their main alternatives are 0.65 for Romania with the US and 0.74 for Morocco with France, while a negative correlation with the main alternative destination only appears for one out of the 61 countries in the sample. The main alternatives are taken from the magnitude of 1990-2000 net migration flows according to the dataset constructed by Docquier, Lowell, and Marfouk (2009).

[42]The sample countries to which the EC Regulation No. 539/2001 applied are Bulgaria, Czech Republic, Hungary, Poland, Romania, Slovakia, and Slovenia.

[43]Similar arguments can be applied to other variables that we control for, such as those referring to the 2004 and 2007 EU enlargements.

[44]Notice that the CCE estimator can accommodate serial correlation and cross-sectional dependence of η_{jkt} in (20) so that the Frees' and Wooldridge' tests are unnecessary. The same observation applies to

With the CCE estimator, we find that a 1.0 percent increase in real GDP per capita leads, after four quarters, to a 3.1 percent reduction in the migration rate to Spain. This effect is only 66 percent of the one estimated in specifications (1)-(2), confirming in this case the expectation that neglecting the influence of the multilateral resistance to migration biases the coefficient of GDP downwards. Similarly, the estimated negative coefficient of the visa requirement is now highly significant, and much larger than the one obtained in the previous specifications. The introduction of a visa requirement for non-immigrant admission to Spain reduces the size of migration flows by 74 percent. This can be compared with the effects from specifications (1)-(2), which pointed to a much smaller (14 percent) and non significant reduction of migration flows. The CCE large estimated effect is in line with the findings on Ecuadorian migration to Spain in Bertoli, Fernández-Huertas Moraga, and Ortega (2011).

These main results are robust to the exclusion of particular countries from the sample, such as OECD, EU-15 countries or, more generally, countries with a GDP per capita higher than the Spanish one. They are also robust to alternative definitions of the dependent variable considering only working age individuals migration or male working age migration. The specifications corresponding to these robustness checks are included in the Appendix B.

6   Concluding remarks

The possible dependence of bilateral migration flows upon the time-varying attractiveness of other destinations represents a source of concern for the econometric analysis of the determinants of migration (Hanson, 2010), as it can introduce an omitted variable bias. This paper has explored the relationship between the stochastic properties of the individual migration decision problem, and the presence of such a bias: when the independence of irrelevant alternatives does not characterize individual migration choices, then bilateral migration rates depend on the opportunities to migrate to other countries, and we labeled this effect Multilateral Resistance to Migration.

Consistent estimates of the determinants of bilateral migration flows can be obtained in the presence of multilateral resistance to migration adopting the Common Correlated Effects estimator proposed by Pesaran (2006). This approach is more general than others proposed in the literature, which either rely on an *ad hoc* controls for the time-varying opportunities to

specification (2).

migrate to other destinations (Mayda, 2010), or require more restrictive assumptions on the stochastic properties of the model and do not allow to identify the effects of origin-specific variables (Ortega and Peri, 2009).

This approach is applied to the analysis of high-frequency Spanish administrative data on bilateral migration flows between 1997 and 2009, which are found to respond quickly and significantly to variations in economic conditions at origin, and to changes in the legal provisions for non-immigrant admission. The econometric analysis shows the empirical relevance of the concern expressed by Hanson (2010) in our data: if not accounted for, multilateral resistance to migration would bias downwards the estimated effect of GDP at origin and upwards the effect of visa policies upon bilateral migration flows to Spain.

References

ANDERSON, J. (2011): "The Gravity Model," *Annual Review in Economics*, 3, 133–160.

ANDERSON, J., AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *The American Economic Review*, 93, 170–192.

——— (2004): "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

BAI, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77(4), 1229–1279.

BALDWIN, R. (2006): "The Euro's Trade Effects," European Central Bank Working Paper Series No. 594, Frankfurt.

BAUM, C. F. (2006): *An Introduction to Modern Econometrics using Stata*. Stata Press.

BEINE, M., F. DOCQUIER, AND C. OZDEN (2011): "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

BERTOLI, S. (2010): "Networks, sorting and self-selection of Ecuadorian migrants," *Annales d'Economie et de Statistique*, 97/98, 261–288.

BERTOLI, S., J. FERNÁNDEZ-HUERTAS MORAGA, AND F. ORTEGA (2010): "Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions," IZA Discussion Paper No. 4957, Bonn.

———— (2011): "Immigration Policies and the Ecuadorian Exodus," *World Bank Economic Review*, 25, 57–76.

BORJAS, G. J. (2003): "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, 118(4), 1335–1374.

BUGAMELLI, M., AND F. PATERNÓ (2009): "Do Workers' Remittances Reduce the Probability of Current Account Reversals?," *World Development*, 37(12), 1821–1838.

CANOVA, F. (2007): *Methods for Applied Macroeconomic Research.* Princeton University Press.

CLARK, X., T. HATTON, AND J. WILLIAMSON (2007): "Explaining U.S. immigration, 1971-1998," *Review of Economics and Statistics*, 89(2), 359–373.

COAKLEY, J., A. FUERTES, AND R. SMITH (2002): "A Principal Components Approach to Cross-Section Dependence in Panels," mimeo, Birbeck College, University of London.

DE HOYOS, R., AND V. SARAFIDIS (2006): "Testing for cross-sectional dependence in panel-data models," *The Stata Journal*, 6(4), 482–496.

DEBREU, G. (1960): "Review of R. D. Luce Individual Choice Behavior," *American Economic Review*, 50(1), 186–188.

DOCQUIER, F., B. L. LOWELL, AND A. MARFOUK (2009): "A gendered assessment of highly skilled emigration," *Population and Development Review*, 35(2), 297–321.

DOCQUIER, F., C. OZDEN, AND G. PERI (2010): "The Wage Effects of Immigration and Emigration," NBER Working Paper No. 16646, National Bureau of Economic Research.

DRISCOLL, J. C., AND A. C. KRAAY (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *Review of Economics and Statistics*, 80(4), 549–560.

DRUKKER, D. M. (2003): "Testing for serial correlation in linear panel-data models," *The Stata Journal*, 3(2), 168–177.

EDIN, P.-A., P. FREDRIKSSON, AND O. ÅSLUND (2003): "Ethnic Enclaves and the Economic Success of Immigrants: Evidence from A Natural Experiment," *Quarterly Journal of Economics*, 118(1), 329–357.

FREES, E. W. (1995): "Assessing cross-sectional correlation in panel data," *Journal of Econometrics*, 69, 393–414.

FRIEDBERG, R. M. (2001): "The Impact of Mass Migration on the Israeli Labor Market," *Quarterly Journal of Economics*, 116(4), 1373–1408.

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

HANSON, G. (2010): "International Migration and the Developing World," in *Handbook of Development Economics*, ed. by D. Rodrik, and M. Rosenzweig, pp. 4363–4414. Amsterdam: North-Holland, Volume 5.

HANSON, G. H., AND C. MCINTOSH (2010a): "Birth Rates and Border Crossings: Latin American Migration to the US, Canada, Spain, and the UK," NBER Working Paper No. 16471, National Bureau of Economic Research.

——— (2010b): "The Great Mexican Emigration," *Review of Economics and Statistics*, 92(4), 798–810.

HANSON, G. H., AND A. SPILIMBERGO (1999): "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the U.S.-Mexico Border," *American Economic Review*, 89(5), 1337–1357.

HAUSMAN, J., AND D. MCFADDEN (1984): "Specification tests for the multinomial logit model," *Econometrica*, 52(5), 1219–1240.

HOECHLE, D. (2007): "Robust standard errors for panel regressions with cross-sectional dependence," *The Stata Journal*, 7(3), 281–312.

IMF (2010a): *International Financial Statistics*. Washington DC.

——— (2010b): *World Economic Outlook - Rebalancing Growth*. April, Washington DC.

INE (2010a): "Estadística de Variaciones Residenciales," Instituto Nacional de Estadística. Madrid. www.ine.es.

———— (2010b): "Padrón Municipal de Habitantes," Instituto Nacional de Estadística. Madrid. www.ine.es.

Lewer, J. J., and H. V. den Berg (2008): "A gravity model of immigration," *Economics Letters*, 99(1), 164–167.

Luce, R. D. (1959): *Individual Choice Behavior*. New York: Wiley.

Mayda, A. M. (2010): "International migration: a panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 23(4), 1249–1274.

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

———— (1978): "Modeling the Choice of Residential Location," in *Spatial interaction theory and planning models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, pp. 75–96. Amsterdam: North-Holland.

———— (2001): "Economic Choices," *American Economic Review*, 91(3), 351–378.

McKenzie, D. J., and H. Rapoport (2010): "Self-selection patterns in Mexico-U.S. migration: The role of migration networks," *Review of Economics and Statistics*, 92(4), 811–821.

Mishra, P. (2007): "Emigration and wages in source countries: Evidence from Mexico," *Journal of Development Economics*, 82(1), 180–199.

Munshi, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics*, 118(2), 549–599.

OECD (2010): *International Migration Outloook - SOPEMI 2010*. Organization for Economic Cooperation and Development.

Orrenius, P. M., and M. Zavodny (2003): "Do Amnesty Programs Reduce Undocumented Immigration? Evidence from IRCA," *Demography*, 40(3), 437–450.

ORTEGA, F., AND G. PERI (2009): "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005," NBER Working Paper No. 14883, National Bureau of Economic Research.

OTTAVIANO, G., AND G. PERI (2010): "Rethinking the Effects of Immigration on Wages," *Journal of the European Economic Association*, forthcoming.

PEDERSEN, P. J., M. PYTLIKOVA, AND N. SMITH (2008): "Selection and network effects–Migration flows into OECD countries 1990-2000," *European Economic Review*, 52(7), 1160–1186.

PESARAN, M. H. (2004): "General Diagnostic Tests for Cross Section Dependence in Panels," IZA Discussion Paper No. 1204, Bonn.

——— (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

PESARAN, M. H., AND E. TOSETTI (2011): "Large panels with common factors and spatial correlation," *Journal of Econometrics*, 161(2), 182–202.

SANTOS SILVA, J. M. C., AND S. TENREYRO (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641–658.

SIMPSON, N. B., AND C. SPARBER (2010): "The Short- and Long-Run Determinants of Unskilled Immigration into U.S. States," Working Paper No. 2010-06, Dept. of Economics, Colgate University.

THEOHARIDES, C., D. MCKENZIE, AND D. YANG (2010): "How Important Are Destination Country Macro Shocks in Determining Migration Flows? Evidence from Filipino Migration," paper presented at the Third Migration and Development Conference, Paris, September 2010.

TRAIN, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

WEN, C.-H., AND F. S. KOPPELMAN (2001): "The generalized nested logit model," *Transportation Research Part B*, 35, 627–641.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

World Bank (2010): *World Development Indicators.* Washington DC.

A   Data sources

A.1   Migration flows

A.1.1   The *Estadística de Variaciones Residenciales*

The EVR is an administrative dataset: municipalities are responsible for keeping the municipal registry up to date and the INE just compiles the information received from the municipalities about variations in the *Padrón*. The EVR registers changes of status in the *Padrón*, both inscriptions and cancelations, with each observation corresponding to one variation. We use the observations referring to the first inscription of foreign-born individuals coming from abroad in the *Padrón* to measure immigration flows to Spain: the EVR contains 6,166,133 of these observations between January 1997 and December 2009.

We can assess the accuracy of the EVR in measuring immigration flows to Spain by comparing it with other possible sources. These alternative sources are represented by the 2001 Population Census, the 2007 *Encuesta Nacional de Immigrantes* (ENI), a special survey for foreign-born individuals, and various rounds of the quarterly Spanish labor survey, *Encuesta de Población Activa* (EPA). The ENI was a special immigrant survey which was only ran once between the last months of 2006 and the first months of 2007, with a sample of approximately 15,000 immigrants.[45] The ENI and the EPA provide information about the year of arrival to Spain of all immigrants, although the EPA does not contain this information for the foreign-born who obtained Spanish citizenship.[46]

A.1.2   Total yearly flows

Our EVR data span the 1997-2009 period. Figure A.1 compares gross immigration flows into Spain according to the EVR with the 2001 Population Census and the ENI,[47] and it shows that the EVR underestimates migration flows before 2000. In January 2000, the Spanish government enacted a new immigration law which included both an amnesty and a provision guaranteeing that immigrants would have access to basic public services such as health and

---

[45]The methodology to locate immigrants was based on past *Padrón* data and it is exactly the same methodology used by the EPA.

[46]This entails that we have to take the EPA numbers as a lower bound, as one out of six immigrants residing in Spain in 2010 was naturalized.

[47]Recall that, the 2007 ENI being a survey, it might fail to enumerate recently arrived immigrants.

Figure A.1: Immigration inflows to Spain (1997-2009)
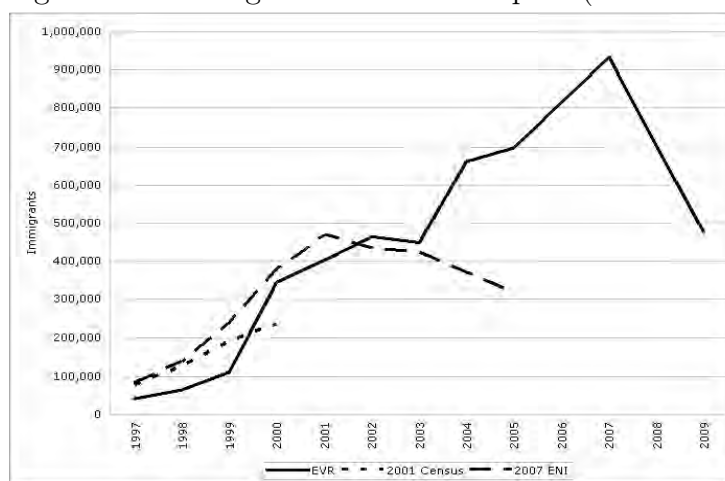
education for their children as long as they register in the *Padrón* (see Bertoli, Fernández-Huertas Moraga, and Ortega (2011)). This shows as a spike in the 2000 EVR data that can be attributed to the registration of both newly arrived immigrants and of those who had come to Spain in the earlier years but had not registered yet.[48] The EVR and ENI series then pretty much coincide for the years 2002 and 2003 but they diverge again for 2004 and 2005 (we do not include the 2006 ENI arrivals because the survey was administered partly in the last months of 2006). There are three possible explanations for the 2004 and 2005 divergence: (*i*) the ENI might be underestimating the number of newly arrived immigrants because of a sampling problem, and of the bias due to the emigration of the foreign-born; (*ii*) the 2005 amnesty may have induced more and more illegal immigrants to register; and (*iii*) the 2004 EU enlargement may have made immigrants from Eastern Europe register massively, even though they may have arrived much earlier.[49]

First, let us consider the likely magnitude of sampling problems in the 2007 ENI. As the methodology to locate immigrants for the ENI is exactly the same methodology used for the EPA, a comparable dataset would be the EPA for the first quarter of 2007, which interviewed around 10,500 immigrants. Figure A.2 shows the implied immigration flows to Spain for different rounds of the first quarter of the EPA between 2006 and 2010. The

---

[48]This may have been also helped by the 2000 and 2001 amnesties, even though being legally in the country is irrelevant for registration.

[49]Incidentally, the 2007 EU enlargement to Romania and Bulgaria is behind the great surge in immigration flows in 2007, as shown in Section 4.1.

Figure A.2: Immigration inflows to Spain (1997-2009)
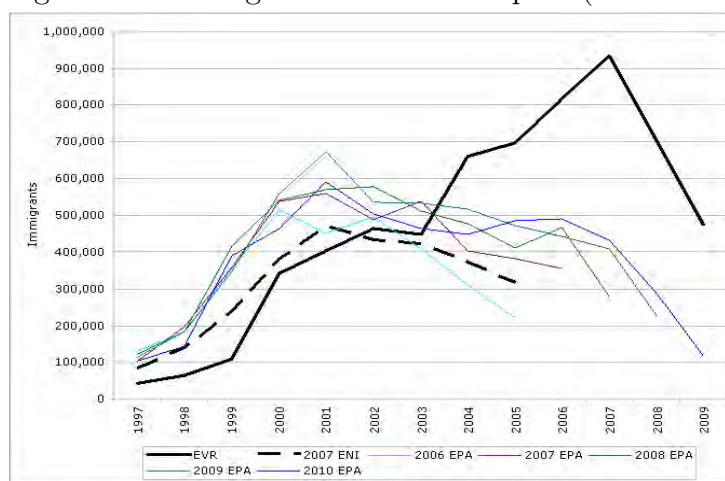
standard errors for the ENI and EPA numbers are between 10,000 and 25,000 immigrants. Hence, what Figure A.2 shows is that sampling problems in the ENI could go on average around half of the way in explaining the difference between the ENI and the EVR. Taking the year 2005, where the discrepancy between the EVR and the ENI is greatest (with 700,000 and 300,000 immigrants respectively), the most recent rounds (2009 and 2010) of the EPA report that 500,000 immigrants arrived.

What about the bias due to the emigration of the foreign-born? The unique data source is represented by the EVR itself, as Figure A.2 shows that there is little hope of gouging the size of return migration or re-migration to third countries from the comparison of different rounds of the EPA. A problem is represented by the fact that it is not mandatory to cancel from the *Padrón* before leaving the country, and the law, while making inscription attractive, does not provide incentives to cancel registrations, and this entails that many episodes of emigration of foreign-born individuals are likely to remain unreported. Fortunately, the law changed in November 2003 (*Ley Orgánica* 14/2003) so that non-EU immigrants (which represented around 60 percent of the total immigrant population in 2005) must renew their inscription every two years, otherwise they are removed from the *Padrón*, with the corresponding variation being recorded in the EVR.[50]

Thus, reliable EVR estimates of the emigration of the foreign-born should be available since 2006. It must be noted though that these figures may be reliable in terms of magnitude

---

[50]The EVR does not provide information on the country of destination of the foreign-born who do cancel from the *Padrón*, and this is why we do not refer to these variations as instances of return migration.

Figure A.3: Migration flows of foreign-born out of Spain (2002-2009)
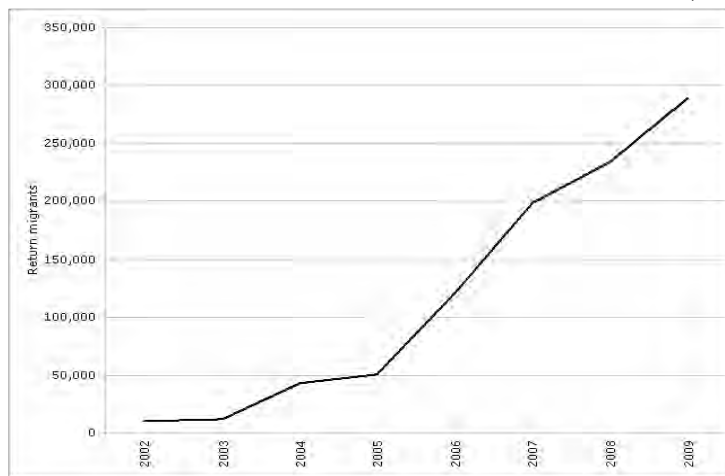
(i.e. every observation corresponds to an instance of emigration of a foreign-born individual), but not necessarily in terms of timing since there could be, at most, a two-year lag between the actual departure and the variation recorded in the EVR. The analysis is further complicated by the fact that the EVR does not provide the information about the date of the first inscription of the foreign-born who cancel from the *Padrón*. Taking all of this into account, Figure A.3 shows the yearly figures of the emigration of the foreign-born according to the EVR.

If we assume, because of the two-year delay, that outflows recorded in 2006 (the first year to which the new law applies) correspond to actual departures in 2004, they would represent 18 percent of the 2004 gross inflow, whereas 2007 outflows correspond to 30 percent of the 2005 gross inflow. Given the uncertainty about the timing of the flows, all that can be said is that emigration could potentially go a large part of the way in explaining the discrepancies between the EVR figures for 2004 and 2005, and the corresponding figures from the ENI. Together with the sampling design problem, emigration of the foreign-born could even go all the way in explaining the observed difference.

With respect to point (*iii*) above, we can safely disregard the role of the 2004 EU enlargement for the 2004 difference in flows. There are two reasons for this: first, none of the enlargement countries accounts for a relevant share of immigration to Spain, with less than 14,000 immigrants in total (2 percent of the 2004 inflow); second, Spain - unlike Sweden, Ireland or the United Kingdom - imposed restrictions on mobility for two years after the enlargement. Thus, the immigrants' situation did not really change until 2006.

Figure A.4: Quarterly Net Migration Flows to Spain
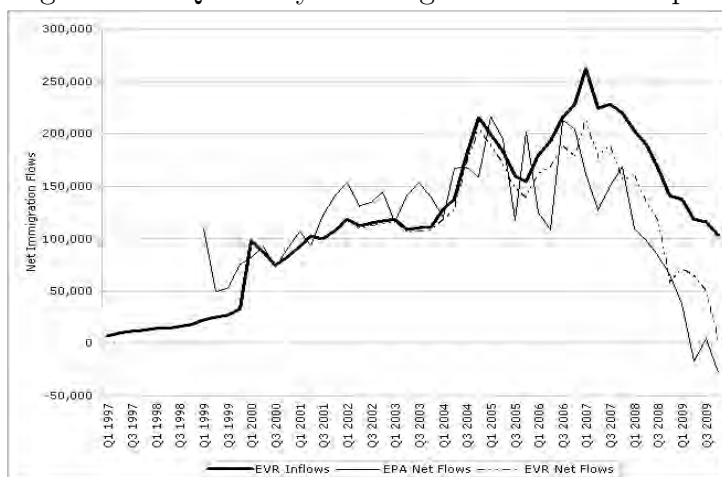
In our empirical analysis, we exploit cross-country high frequency variations in migration flows so that all of the discussed measurement problems in the EVR end up being absorbed by our quarter fixed effects. Measurement problems related to particular countries of origin would only be an issue as long as they may not be absorbed as well by our origin-year fixed effects.

A.1.3   Total quarterly flows

The EPA represents the only other data source for which variation on migration flows at the quarterly level can be obtained. By subtracting the stock of migrants in a given quarter from the stock of migrants in the following quarter we can obtain a measure of *net* migration flows. The quarterly migration series that we produce with this methodology can be compared to the net migration flows obtained from the EVR (recall that figures for the emigration of the foreign-born can only be considered reliable after 2006). This is what is done in Figure A.4.

The comparison of both time series indicates that the general trend and magnitude of the flows is highly comparable in the two sources, especially taking into account the standard errors associated with the EPA net flow. The raw correlation between the two net flows series is 0.78, which is extremely high considering the uncertainty involving the timing of emigration flows from the EVR.

43

A.1.4 Immigration flows by origin

The EVR and the ENI look much more alike when we move to a country-level analysis. The correlation coefficient between origin-year observations from the EVR and from the ENI between 1997 and 2005 is 0.88 (calculated over 604 country-year observations). If we restrict the ENI sample to those country-year pairs for which there were at least 10 observations, we are left with 177 country-year observations, for which the correlation with the EVR is still 0.84.[51]

We run basic regressions to check to what extent ENI origin-year observations can explain EVR origin-observations: when we did so for the 177 common observations for which there were at least 10 individuals in the ENI sample, the result is a coefficient reassuringly equal to 1.00.[52] This could hide differences on a country by country basis but, when we run origin-specific regressions (with the caveat that the highest number of observations is 9 in these regressions), we could not reject the coefficient on the ENI numbers being 1 for any country but Colombia at a 95 percent confidence level (with a p-value of 0.0497).

When we run a regression with origin fixed effects, the resulting coefficient was 0.97 (not statistically different from 1 at a 99 percent confidence level). We also run a regression with time fixed effects exploiting the cross-sectional variation in the data, and the estimated coefficient was again 1.00. However, the last specification shows that the 2004 and 2005 year fixed effects are significant, which indirectly suggests that point (*iii*) above about the 2005 amnesty (announced in the last months of 2004) did play a relevant role.

A.2 Spanish immigration policies

We detail below how each of the ten variables that describe how Spanish immigration policies changed over our period of analysis (1997-2009) were built, and the corresponding legal sources.

January 2000 Amnesty - the dummy variable takes value 1 for all countries from January 2000 (source: *Ley Orgánica* 4/2000).

---

[51]The share of total immigrants covered by this restriction is 85 percent both in the ENI and in the EVR.

[52]Of course, this result hides differences on a year by year basis: running yearly regressions, we obtain a coefficient around 0.3 for the years 1997-1999, around 1 for 2000-2003, 1.3 in 2004 and 2.3 in 2005.

November 2004 Amnesty - the dummy variable takes value 1 for all countries from November 2004 (source: *Real Decreto* 2393/2004).

EU-15 - the dummy variable takes value 1 if the country of origin belongs to the European Union as of 1997. Thus, it is 1 for Austria, Belgium, Denmark, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, United Kingdom, Germany and Sweden (source: www.europa.eu.int).

Schengen Area - the dummy variable takes value 1 from the inclusion of a country in the Schengen Area. It is 1 in the whole sample period for Belgium, France, Luxembourg, Monaco, Netherlands, Portugal and Germany; 1 from November 1997 for Italy, San Marino and the Holy See; 1 from December 1997 for Austria; 1 from April 2000 for Greece; 1 from April 2001 for Denmark, Finland, Iceland, Norway and Sweden; 1 from April 2008 for Hungary, Malta, Poland, Latvia, Estonia, Lithuania, the Czech Republic, Slovakia and Slovenia; and 1 from January 2009 for Switzerland (source: www.europa.eu.int).

EU May 2004 Eastern Enlargement - the dummy variable takes value 1 from May 2004 for Cyprus, Hungary, Malta, Poland, Latvia, Estonia, Lithuania, Czech Republic, Slovakia and Slovenia (source: www.europa.eu.int).

EU January 2007 Romania and Bulgaria Enlargement - the dummy variable takes value 1 from January 2007 for Romania and Bulgaria (source: www.europa.eu.int).

Visa requirement for non-immigrant admission - the dummy variable takes value 1 for those countries and periods for which a visa was required to enter Spain. It is 1 for all values with the exception of the following: members of the EU-15 group; Andorra; Iceland; Norway; Liechtenstein; Croatia; country-month pairs for which the Schengen area dummy is 1; Eastern Enlargement (2004 and 2007) countries plus Switzerland from April 2001; Chile; Peru; Argentina; Bolivia until March 2007; Colombia until December 2001; Ecuador until July 2003; Venezuela; Paraguay; Brazil; Uruguay; Mexico; Costa Rica; El Salvador; Guatemala; Honduras; Panama; Nicaragua; Australia; New Zealand; Canada; United States; South Korea; Brunei; Israel; Japan; Malaysia; Singapore; Antigua and Barbuda, Bahamas, Barbados, Saint Kitts and Nevis, Seychelles and Mauritius from June 2009. The sources are Schengen Area regulations (www.europa.eu.int; www.maec.es and www.boe.es) and bilateral agreements of Spain with Latin American countries (www.mtin.es and www.boe.es).

Bilateral agreement on double nationality - the dummy takes value 1 if a bilateral agreement on double nationality with Spain exists: Costa Rica, Guatemala, Honduras,

Nicaragua, Dominican Republic, Argentina, Bolivia, Colombia, Chile, Ecuador, Paraguay and Peru (sources: www.mtin.es and www.boe.es).

Bilateral agreement on social security - the dummy takes value 1 if a bilateral agreement on Social Security with Spain exists: Mexico, Argentina, Brazil, Chile, Ecuador, Paraguay, Peru, Uruguay and Venezuela for the whole period; Panama until May 2000; Dominican Republic from July 2006; and Colombia from March 2008 (sources: www.mtin.es and www.boe.es).

Bilateral agreement on contracts at origin - the dummy takes value 1 from the moment when a bilateral agreement that allows to sign in the country of origin a labor contract with a Spanish employer is applied: Colombia from June 2001, Ecuador from July 2001, Dominican Republic from February 2002 and Peru from August 2004 (sources: www.mtin.es and www.boe.es).

A.3   GDP data

We gathered real GDP quarterly data for all countries of origin with a positive total number of immigrants in all the quarters, and we were able to find these data for 61 origin countries, representing 87 percent of total immigration flows to Spain between 1997 and 2009.[53] The main data source was represented by IMF (2010a), which we combined with data from IMF (2010b) and from various Central Banks.[54] When the original series of real quarterly GDP data were not seasonally adjusted, we implemented the adjustment regressing the log of real GDP on a linear time trend and quarterly dummies, as suggested by Baum (2006).

---

[53]The population residing in these countries amount to 51 percent of the world total.

[54]IMF (2010b) provides information on the rate of growth of real quarterly GDP for several countries and regional aggregates; for each origin country, Table A.1 reports whether the figures from IMF (2010b) are country- or region-specific.

Table A.1: Data sources for quarterly real GDP

Country	Source	from	to	SA	obs.
Argentina	IFS	1997q1	2009q4	no	52
Australia	IFS	1997q1	2009q4	yes	52
Austria	IFS	1997q1	2009q4	no	52
Belgium	IFS	1997q1	2009q4	yes	52
Bolivia	IFS	1997q1	2009q3	no	51
	WEO, LAC	2009q4	2009q4	no	1
Brazil	IFS	1997q1	2009q4	no	52
Bulgaria	WEO, EE	1999q1	2001q4	no	12
	IFS	2002q1	2009q4	no	32
Canada	IFS	1997q1	2009q4	yes	52
Chile	IFS	1999q1	2002q4	no	16
	WEO, LAC	2003q1	2009q4	no	28
Colombia	IFS	1999q1	1999q4	yes	4
	IFS	2000q1	2009q4	yes	40
Costa Rica	WEO, LAC	1999q1	1999q4	yes	4
	IFS	2000q1	2009q4	no	40
Croatia	IFS	1997q1	2009q4	no	52
Czech Republic	IFS	1997q1	2009q4	no	52
Denmark	IFS	1997q1	2009q4	no	52
Dom. Republic	Central Bank	1997q1	2009q4	no	52
Ecuador	IFS	1997q1	2007q3	yes	43
	Central Bank	2007q4	2009q4	yes	9
Egypt	WEO, MENA	1999q1	2001q4	yes	12
	IFS	2001q1	2009q4	no	32
El Salvador	WEO, LAC	1997q1	2005q4	yes	36
	IFS	2006q1	2008q1	yes	9
	WEO, LAC	2008q2q1	2009q4	yes	7
Finland	IFS	1997q1	2009q4	no	52
France	IFS	1997q1	2009q4	yes	52

(continued)

Table A.1: Data sources for quarterly real GDP (continued)

Country	Source	from	to	SA	obs.
Georgia	WEO, CIS	1999q1	2002q4	yes	16
	IFS	2003q1	2009q4	no	28
Germany	IFS	1997q1	2009q4	yes	52
Greece	WEO, Euro	1999q1	2000q4	yes	8
	IFS	2001q1	2009q4	no	36
Guatemala	WEO, LAC	1999q1	2000q4	no	8
	Central Bank	2001q1	2009q4	yes	36
Hungary	IFS	1997q1	2009q4	no	52
Iceland	IFS	1997q1	2009q4	no	52
India	WEO, country	1999q1	2006q4	no	32
	IFS	2007q1	2009q4	no	12
Indonesia	IFS	1997q1	2009q4	no	52
Iran	IFS	1997q1	2007q4	no	44
	WEO, Emerg.	2008q1	2009q4	no	8
Ireland	IFS	1997q1	2009q4	no	52
Israel	IFS	1997q1	2009q4	no	52
Italy	IFS	1997q1	2009q4	yes	52
Japan	IFS	1997q1	2009q4	yes	52
Jordan	IFS	1997q1	2009q4	no	52
Luxembourg	IFS	1997q1	2009q4	no	52
Mexico	WEO, LAC	1999q1	2002q4	no	16
	IFS	1997q1	2003q1	yes	28
Morocco	WEO, MENA	1999q1	2004q4	yes	24
	IFS	2005q1	2009q4	yes	20
Netherlands	IFS	1997q1	2009q4	yes	52
Nicaragua	WEO, LAC	1999q1	2002q4	no	16
	Central Bank	2003q1	2009q4	no	28
Norway	IFS	1997q1	2009q4	no	52
Panama	WEO, LAC	1999q1	2003q2	yes	18
	IFS	2003q3	2006q1	no	11
	WEO, LAC	2006q2	2009q4	yes	15
Paraguay	WEO, LAC	1999q1	2005q4	yes	28
	IFS	2006q1	2008q3	no	11
	WEO, LAC	2008q4	2009q4	yes	5

(continued)

Table A.1: Data sources for quarterly real GDP (continued)

Country	Source	from	to	SA	obs.
Peru	IFS	1997q1	2009q4	no	52
Philippines	IFS	1997q1	2009q4	no	52
Poland	IFS	1997q1	2009q4	no	52
Portugal	IFS	1997q1	2009q4	no	52
Romania	IFS	1997q1	2009q4	no	52
Russia	IFS	1997q1	2009q3	no	51
	WEO, country	2009q4	2009q4	no	1
Slovakia	IFS	1997q1	2009q4	no	52
Slovenia	IFS	1997q1	2009q4	no	52
South Africa	IFS	1997q1	2009q4	no	52
South Korea	IFS	1997q1	2009q4	no	52
Sweden	IFS	1997q1	2009q4	no	52
Switzerland	IFS	1997q1	2009q4	yes	52
Turkey	IFS	1997q1	2009q4	no	52
Tunisia	WEO, MENA	1999q1	2000q4	yes	8
	IFS	2001q1	2007q4	no	28
	WEO, MENA	2008q1	2009q4	yes	8
Ukraine	WEO, EE	1999q1	2000q4	no	8
	IFS	2001q1	2009q4	yes	36
United Kingdom	IFS	1997q1	2009q4	yes	52
Uruguay	Central Bank	1997q1	2008q4	no	48
	WEO, LAC	2009q1	2009q4	no	4
USA	IFS	1997q1	2009q4	yes	52
Venezuela	Central Bank	1997q1	2009q4	yes	52

Notes: SA describes whether the original series was seasonally adjusted.

B Robustness checks

Table B.1: Robustness checks on the determinants of migration

Dependent variable: log of quarterly emigration rate

Specification		Baseline		No OECD		No EU-15	
Estimation methods		FE	CCE	FE	CCE	FE	CCE
Regressors	Lags						
Log real GDP per capita	1	-1.57	-1.57	-1.72	-1.59	-1.68	-1.72
		[0.24]***	[0.29]***	[0.29]***	[0.35]***	[0.27]***	[0.33]***
	2	-1.05	-0.46	-1.22	-0.17	-1.05	-0.43
		[0.26]***	[0.29]	[0.32]***	[0.34]	[0.29]***	[0.33]
	3	-0.93	-0.60	-0.54	-0.21	-1.05	-0.79
		[0.27]***	[0.28]**	[0.32]*	[0.34]	[0.30]***	[0.33]**
	4	-1.18	-0.52	-1.17	-0.06	-1.42	-0.66
		[0.26]***	[0.31]*	[0.32]***	[0.32]	[0.29]***	[0.36]*
GDP per capita,		-4.74	-3.14	-4.65	-2.03	-5.20	-3.59
cumulated effect		[0.37]***	[0.53]***	[0.49]***	[0.67]***	[0.43]***	[0.62]***
Visa requirement	0	-0.15	-1.34	-0.50	-1.32	-0.14	-1.34
		[0.13]	[0.30]***	[0.20]**	[0.32]***	[0.15]	[0.34]***
Other policy controls		yes	yes	yes	yes	yes	yes
Quarter fixed effects		yes	yes	yes	yes	yes	yes
Origin-year fixed effects		yes	yes	yes	yes	yes	yes
Observations		2,776	2,776	1,504	1,504	2,120	2,120
Countries of origin		61	61	34	34	47	47
Frees' test		2.71	-	1.63	-	1.80	-
(p-value)		(0.00)	-	(0.00)	-	(0.00)	-
Wooldridge's test		32.64	-	18.86	-	22.17	-
(p-value)		(0.00)	-	(0.00)	-	(0.00)	-
Cross-sectional averages		-	2.11	-	2.07	-	1.88
(p-value)		-	(0.00)	-	(0.00)	-	(0.00)

Notes: standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1; observations are weighted by population at origin; the baseline specification follows table 2; No OECD drops high-income OECD countries; No EU-15 drops the members of the European Union before the 2004 enlargement.

Table B.1: Robustness checks on the determinants of migration (continued)

Dependent variable: log of quarterly emigration rate

Specification		Lower GDP		Working Age		Male Working Age	
Estimation methods		FE	CCE	FE	CCE	FE	CCE
Regressors	*Lags*						
Log real GDP per capita	1	-1.63	-1.51	-1.38	-1.32	-1.23	-0.92
		[0.27]***	[0.33]***	[0.25]***	[0.29]***	[0.32]***	[0.39]**
	2	-0.94	-0.05	-1.29	-0.54	-2.15	-0.75
		[0.29]***	[0.32]	[0.26]***	[0.29]*	[0.33]***	[0.40]*
	3	-0.83	-0.63	-1.04	-0.82	-0.30	0.08
		[0.30]***	[0.32]**	[0.27]***	[0.28]***	[0.34]	[0.38]
	4	-1.19	-0.10	-1.46	-0.57	-2.49	-1.62
		[0.30]***	[0.36]	[0.26]***	[0.31]*	[0.32]***	[0.43]***
GDP per capita,		-4.59	-2.29	-5.16	-3.25	-6.17	-3.21
cumulated effect		[0.46]***	[0.63]***	[0.37]***	[0.53]***	[0.47]***	[0.72]***
Visa requirement	0	-0.14	-1.30	-0.11	-1.18	-0.09	-1.15
		[0.15]	[0.31]***	[0.13]	[0.29]***	[0.17]	[0.36]***
Other policy controls		yes	yes	yes	yes	yes	yes
Quarter fixed effects		yes	yes	yes	yes	yes	yes
Origin-year fixed effects		yes	yes	yes	yes	yes	yes
Observations		1,872	1,872	2,776	2,776	2,765	2,765
Countries of origin		42	42	61	61	61	61
Frees' test		1.88	-	2.87	-	2.77	-
(p-value)		(0.00)	-	(0.00)	-	(0.00)	-
Wooldridge's test		23.02	-	63.51	-	17.15	-
(p-value)		(0.00)	-	(0.00)	-	(0.00)	-
Cross-sectional averages		-	1.95	-	2.20	-	2.06
(p-value)		-	(0.00)	-	(0.00)	-	(0.00)

Notes: standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1; observations are weighted by population at origin; specification Lower GDP drops all countries with a GDP per capita higher than the Spanish one in PPP terms in 1999 ((World Bank, 2010)); specification Working Age computes the dependent variable only on working age individuals (between 16 and 65) when they enter Spain; specification Working Age Male further restricts the definition of the dependent variable to working age males.