IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11143

Immigrant Locations and Native Residential Preferences: Emerging Ghettos or New Communities?

Jesús Fernández-Huertas Moraga
Ada Ferrer-i-Carbonell
Albert Saiz

NOVEMBER 2017

IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11143

Immigrant Locations and Native Residential Preferences: Emerging Ghettos or New Communities?

Jesús Fernández-Huertas Moraga
Universidad Carlos III de Madrid and IZA

Ada Ferrer-i-Carbonell
IAE, CSIC, Barcelona GSE and IZA

Albert Saiz
Massachusetts Institute of Technology and IZA

NOVEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Immigrant Locations and Native Residential Preferences: Emerging Ghettos or New Communities?*

While the impact of immigrants on labor markets may be small, strong political movements voicing opposition to the growth of resident foreign-born populations are on the upswing. We study whether natives voted with their feet in reaction to the largest and fastest migration shock in the OECD. The inflow, causing the population of Spain to grow by 10 percent between 1998 and 2008, represented largely a new phenomenon the size of which had not been factored into previous expectations, thereby providing quasi-experimental sources of variance. Our results show that immigrant inflows caused mild native flight from denser, established neighborhoods, but also more real estate development there. In parallel, both natives and immigrants were concurrently moving into new booming suburban communities, resulting in no changes in overall measures of ethnic segregation. In contexts where large ethnic minority arrivals spur the creation of new neighborhoods, conventional empirical methods may overstate the degree of segregationist behavior.

JEL Classification: F22, J61, D33

Keywords: international migration, residential segregation, white flight

Corresponding author:
Jesús Fernández-Huertas Moraga
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe, Madrid
Spain
E-mail: jesferna@eco.uc3m.es

* We are grateful to Simone Bertoli, Amparo González, Joan Monràs, Rosella Nicolini and Francesc Ortega for their comments and suggestions. We also thank seminar participants at the Universidad de Alicante, IREA-UB Institute (Universitat de Barcelona), IEB (Universitat de Barcelona), the Department of Applied Economics (Universitat Autònoma de Barcelona), CERDI (Université d'Auvergne), IRES (Université Catholique de Louvain), IZA, UC Riverside, Williams College, Trinity College Dublin, University of British Columbia, the 7th INSIDE Workshop at the Barcelona GSE Summer Forum, the 4th CEPII-OECD Conference on \Immigration to OECD Countries" in Paris and the CEMIR Junior Economist Workshop on Migration Research in Munich. We also wish to thank Lídia Brun for her excellent work of data collection and Blake Willmarth and Erdal Karahan for their research assistance. We acknowledge the financial support received by the RecerCaixa grant. The usual disclaimers apply.

1 Introduction

International migration has become one of the central political issues of our time. The share of immigrants over the total population of high-income OECD countries increased by approximately 1 percentage point every five years between 1990 and 2010: from 7 percent to 11 percent (World Bank, 2013). This has led economists to investigate how immigrants affect destination economies. Most studies —focusing on labor markets— arguably show that wage impacts on natives are relatively small.[1]

However, attitudes towards immigrants depend on other issues, such as their impact on welfare benefits or ethnic identities (Dustmann and Preston, 2007). Some authors, like Borjas (2013) or Saiz and Wachter (2011), claim that the larger impacts of immigration may not be found in the labor market but rather in other interaction spaces between natives and immigrants. After all, robotization, capital deepening, Information Technologies, Artificial Intelligence, trade, global outsourcing, and off-shoring, have also profound impacts on employment, wages, and the labor market, but do not seem to galvanize opinion in the same way as immigration. This suggests that the residential aspects of this phenomenon are perceived as critically important by natives. In this context, we need to study where immigrants settle, and how their settlement decisions affect natives as revealed by the residential choices of the latter.

To this end, we focus on one of the world's largest and swiftest immigration experiences. In only ten years, between 1998 and 2008, the immigrant share in Spain increased from 3 percent to a staggering 13 percent of the population. We perform a comprehensive study of natives' residential responses, combining microdata on exact addresses of the resident population, before and after the immigration shock, with distance to amenities and socioeconomic characteristics of neighborhoods. To our knowledge, this is the most complete set of neighborhood characteristics as controls in the ethnic segregation literature. Therefore, our results are less likely to be contaminated by omitted neighborhood characteristics than in previous work. Building on Saiz and Wachter (2011) and Kasy (2015), we combine the large set of controls with instrumental variables (IV) based on geographic diffusion of immigrant inflows, in order to get an estimate of the causal impact of local immigration shocks on

[1]See Docquier, Ozden, and Peri (2014), for example. Some authors argue that the distributional consequences for particular groups can be very large (Borjas and Monras, 2016) although others disagree (Clemens and Hunt, 2017).

natives' outflows.

In an environment of massive and ex novo inflows, we show that: (i) immigrants displaced natives from city centers and dense suburban areas at a mild rate of three to one: one native fleeing for every three immigrant arrivals; and (ii) residential construction also increased in the central areas with more immigrant arrivals. We argue that these two facts ares inconsistent with the existence of strong discriminatory residential behavior against migrants. We also show that: (iii) new housing developments in suburban areas saw both immigrant and native arrivals; and (iv) the resulting effect on average immigrant segregation was neutral. The results point to interesting dynamics that are not captured by conventional tipping and segregation models, which assume fixed neighborhood sets and often ignore the role of real estate development and the creation of new neighborhoods.

Most of the literature on the effect of ethnic displacement has centered in the US (Cutler, Glaeser, and Vigdor, 2008; Card, Mas, and Rothstein, 2008; Saiz and Wachter, 2011) where migrant inflows of ethnic minorities to cities have been progressive and continuous. The richness of our dataset is similar to Scandinavian ones, though, where research has been conducted with regards to smaller inflows in countries that had been also gradual net recipients in recent history (Edin, Fredriksson, and Åslund, 2003; Damm, 2009; Jofre-Monseny, Dahlberg, and Fredriksson, 2012). Beyond the US and Scandinavia, a number of papers have been written on other countries such as the UK (Giulietti, 2009; Bell and Machin, 2013; Hatton and Tani, 2005), Italy (Boeri, De Philippis, Patacchini, and Pellizzari, 2014; Brücker, Fachin, and Venturini, 2011; Accetturo, Manaresi, Mocetti, and Olivieri, 2014), the Netherlands (Gautier, Siegmann, and Van Vuuren, 2009), or France (Verdugo, 2011). These studies have been conducted at different levels of geographical aggregation and using alternate empirical methods.

Our geocoded data allow us to study the phenomenon in a context where its magnitude and speed were very sizable and unpredicted. This is a substantially important global migration episode worth understanding well. In addition, we are fortunate to have very detailed micro data on immigrant and native locations, and a much broader coverage of the neighborhood characteristics that could have impacted native mobility even in the absence of migratory flows. The richness of the data also allows us to thread carefully with causal inference, to investigate potential non-linearities in the magnitude of "native flight," to study margins of adjustment in construction and residential densities, and to analyze the spatial dimension of the relevant phenomenon.

There is a number of papers dealing with the measurement of residential segregation in Spain. For example, Pareja-Eastaway (2009) and García-López (2012) measure segregation in Barcelona, while Blázquez, Llano, and Moral (2010) relate commuting times to segregation (1996-2004), and Echazarra (2010) studies the correlates of neighborhood segregation (2001-2006) in Madrid. Bosch, Carnero, and Farré (2010) study the existence of ethnic discrimination against immigrants in the rental market. Ballester and Vorsatz (2014) produce cross-sectional segregation measures.[2] Finally, Maza, Villaverde, and Hierro (2013) look at the determinants of immigrants' location across Spanish provinces. While providing important and necessary measurements, none of these studies analyze the causal reaction of natives to localized immigrant arrivals.

The immigration boom in this country coincided with the housing boom, we suspect not unrelatedly. Nicodemo and Raya (2012) study the distribution of changes in house prices between 2004 and 2007. González and Ortega (2013) use inter-regional variation to find that *'immigration was responsible for one quarter of the increase in prices and about half of the construction activity over the decade'* 2000-2010. We focus on neighborhood-level variation, where comprehensive data on prices during that period do not exist. Our estimates of the effect of immigrant arrivals on native mobility are therefore reduced-form, comprising both: (i) native preferences for neighborhood ethnic composition; and (ii) equilibrium effects on housing prices and native price-elasticities of substitution between neighborhoods. Nevertheless, we do exploit data on residential housing construction, which turns to be a critical issue. Studies in the ethnic segregation literature that abstract from housing supply effects, for instance by chiefly focusing on changes in racial shares, may therefore be lacking. We argue that, in rapidly growing cities, we need to study relative demographic inflows and new housing construction concurrently.

The next section describes and benchmarks the immigration boom and our main data sources. In Section 3, we sketch our methodology by introducing our empirical model and contribution. We also explain the construction of our dataset and show summary statistics. Section 4 presents our main empirical estimates and relates them to the emergence of new metropolitan suburbs. Section 5 concludes by discussing how our findings can shed light on

[2]Without focusing on immigrants, De la Roca and Puga (2017) and De la Roca (2017) study the sorting of individuals by skills and city size.

the current literature on “white flight” and segregation.

2 Largest and Fastest Immigration Boom in Recent History

Spain’s experience in the early 2000s offers a unique opportunity to study a quasi-exogenous, unexpected, and economically meaningful immigrant shock relative to the magnitude of the initial population.[3] Figure 1 compares the evolution of the share of foreign-born individuals living in the country versus other OECD countries between 1990 and 2010. The share of foreign-born individuals in 1990 was 2.1 percent. By 2010, it had reached 13.8 percent, more than a 6-fold increase.[4] Using the official population registry (which we describe later) and looking at total numbers (Table 1), only the United States received more immigrants during this period (19.6 million versus 5.5 million), and the third ranking country was Germany, with 4.8 million. However, American and German inflows were much smaller compared to their total populations, and both countries were starting from much higher foreign-born shares.

(Figure 1) (Table 1)

The composition of the immigrant population also changed during this period (Table 2), with Latin American countries, mainly Ecuador, Colombia and Argentina, becoming the main origins. Another growing area of origin was Eastern Europe, mostly Romania, which consolidated a 19.2 percent of the immigrant population. 12.3 percent arrived from the Maghreb, chiefly Morocco. Other smaller inflows arrived from developing countries in Asia (China, Pakistan, Philippines, and India), and Subsaharan Africa. Conversely, the share of most developed countries in the foreign-born population went down to 22.3 percent, from 47.7 percent in 1998. This change in the composition of the immigration origins provide more strength to the use of our Card-like instruments (Card, 2001) to establish causality

[3] The closest comparison is to be found in Israel in the early 1990s (Friedberg, 2001).

[4] Cambodia’s share raised from 0.4 to 2.4 and is second in the world with a 490 percent increase, which is still notably smaller than the Spanish 548 percent increase. The closest experiences among developed countries are Finland (232 percent increase, from 1.3 to 4.2), Iceland (211 percent increase, from 3.8 to 11.7) and Ireland (210 percent increase, from 6.5 to 20). If we look at the total number of immigrants, the Spanish increase (669 percent) was the second in the world only after Cambodia’s (775 percent). Own calculations from World Bank (2013).

(Ruist, Stuhler, and Jaeger, 2017).

(Table 2)

Figure 2 shows the share of immigrants by province in 2001 and 2008. While immigrant concentration increased substantially in all provinces, Madrid, the Canary Islands, and the Mediterranean provinces accounted for three quarters of the inflows (compared to their 53.3 percent share of the native population).

(Figure 2)

The country's 83 metropolitan areas —as defined by the Ministry of Housing (Ministerio de Vivienda, 2007)— received 72.7 percent of the foreign population inflows, compared to their 66.9 percent share of the native population. Administratively, these 83 metropolitan comprise 744 municipalities, on which we focus in this work; the remaining 7,368 municipalities are small, only 70 of them having more than 25,000 inhabitants.

Following Massey and Denton (1988), we start by calculating dissimilarity and isolation indices. Both measures compare the concentration of minorities to the distribution of the native population across neighborhoods. As in Cutler, Glaeser, and Vigdor (2008), we define minority groups by country of birth.[5] Neighborhoods are defined as censal sections fixed in 2008 —which we refer to as census tracts hereinafter—, an administrative division used for electoral purposes. With some exceptions, each census tract was designed to have between 500 and 2,500 inhabitants. There were 21,505 such neighborhoods in metropolitan areas in 2008 (35,586 in the whole country). Figure 3 shows the population-weighted average of national isolation indexes across foreign national groups. For comparison purposes, the same indices as calculated by Cutler, Glaeser, and Vigdor (2008) are shown at the same scale. The average dissimilarity coefficient slightly decreased to around a value of 0.45, while the mean isolation index very slightly increased during the period. This situates the dissimilarity-based segregation of immigrants at the same levels as US immigrants in the 1970s, and their isolation was lower than the minimum ever attained by US immigrants (slightly below 0.04 in 1950).

(Figure 3)

Overall segregation indexes hide a vast range of experiences by metro area and nation-

[5]The interpretation of the dissimilarity index is the share of the minority group members in a metro area that should change residency in order to be distributed across neighborhoods in the same proportions as natives. The isolation index measures the probability that an individual from a group will meet at random with another group member in her neighborhood.

ality (Fernández-Huertas Moraga, Ferrer, and Saiz, 2011), but their evolution is robust to different definitions of the neighborhood, groupings of nationalities, or to using more sophisticated segregation indices (Ballester and Vorsatz, 2014). Such results are surprising: due to the large nature of the inflows, immigration was perceived as a major social issue among the autochthonous population. 48 percent of respondents to the government's periodical sociological poll in 2008 deemed the level of immigration "excessive" and a further 31 percent thought is was "elevated," with only 19 percent thinking that immigration levels were "acceptable" (Cea D'Anconna and Valles Martínez, 2014). The combination of apparent social preoccupation with an actual lack of segregationist behavior begets the question: what were the mechanisms and flows of migrants and natives that combined into such uneventful aggregate results?

3 Methodology

3.1 Data and Geographies

During most of this period, in particular since 1998-1999, each January 1st the national statistics office (INE, Instituto Nacional de Estadística) gathered the full picture of the location of every person living in the country through the "Padrón Municipal de Habitantes" (INE, 2009). The Padrón is the official registry of population in each municipality.[6] Individuals had strong incentives to register where they live upon moving to a new residence. Registration allowed them to enjoy municipal services and granted them access to regionally-provided ones, such as education and health. The principal criterium by which residents during the period were assigned to schools and hospitals was residential proximity, as measured by their official registration. In fact, if not registered, they had no access to them. All newborn children were immediately registered before discharge, and deceased persons removed upon death certification: this dataset should include as close to 100 percent of the country's population as possible.

Immigrants faced additional incentives to register as soon as they arrived in the country.

[6]Local public funding levels are determined by this official measurement of population. Hence municipalities may have incentives to inflate these numbers while the central government has the opposite ones. Overall, Fernández-Huertas Moraga, Ferrer, and Saiz (2011) estimate that the Padrón inflated the actual population numbers by about 2 percent in 2001.

Registration had been used to prove residency in the periodical regularizations since the Law 4/2000. Hundreds of thousands of immigrants took advantage of being duly registered in the 2000, 2001 and 2005 amnesties. Also, the Law 4/2000 specifically provided access to all public services to immigrants who registered, including the undocumented. Table 1 shows the evolution of the total population, the foreign-born, and their share between 1998 and 2008 according to the Padrón. Population increased by 16 percent between 1998 and 2008. Immigration was directly responsible for 77 percent of such growth.[7]

We have now obtained access to individual micro-data from the official registry during the immigration boom. We have therefore the ability to determine the location and some characteristics of every single one of the approximately 45 million residents each year. Our Padrón data includes the exact address of every registered individual living in the country between 1999 and 2008, their gender, date of birth, place of birth (municipality for natives, or country of birth for immigrants), education level (only up to high school), and nationality. Having the exact localization of each individual, which we geocoded at each unique address, allows us to circumvent the problems generated by changes in administrative neighborhood definitions. The data are anonymized and —unfortunately for the researcher, but due to confidentiality concerns— do not contain household identifiers (so we do not know which persons live in the same apartment) or individual IDs (which prevents us from following individuals after they change their address).

We complement this unique population dataset with two main additional sources: socioeconomic variables from the 2001 Census (INE, 2005) and amenity locations from ESRI StreetMap Premium Europe NAVTEQ 2009 Release 2 (ESRI, 2009). Socioeconomic variables are available at the 2001 census tract level, to which we matched each of the about 7 million unique addresses in the registry. As for amenities, we calculated the distance of every address in our dataset to a series of 62 different points of interest (POIs), such as hospitals, highway exits, schools, bus and subway stops, ATMs, etc (see Table 3 for a full itemization). For all the separate addresses in the national registry and for each of the 62

[7]Padrón figures are not completely reliable for immigrants until 2001, as documented by Bertoli and Fernández-Huertas Moraga (2013). Before the Law of 4/2000, municipalities could use some discretion regarding the registration of immigrants, particularly if they were undocumented. This means that some of the registrations in 2001 correspond to earlier arrivals, likely from the two previous years. To avoid measurement error in the exact timing of registration we use long differences (2001-2008) in population stocks as a proxy for flows in our study.

features we constructed measures of access to amenities. One such measure is the distance to the closest POI in the feature list. For example, we have the distance between each address and its closest hospital. The others are gravity-style measures: here, we focus on the sum of all POIs in the province weighted by the inverse of the square of the distance between the POI and the address.[8]

While we identify the 2001-defined census tract of each unique address, we prefer to create and use exogenous geographic definitions. A substantial proportion of the interesting dynamics of ethnic segregation in growing cities naturally occurs in areas that had been scarcely populated earlier, but which saw new residential development concurrently with migrant arrivals. Pre-determined administrative boundaries lump together these newly-developed areas with pre-existing denser settlements. In addition, government or census-defined neighborhood boundaries are typically defined ex post in order to configure relatively homogeneous demographic and socio-economic units, or in ways that could be endogenous to the phenomena of interest. For instance, census tracts with a larger share of non-voting immigrants are larger. More generally, the use of official neighborhood definitions —as in most of the extant literature— may entail measurements of segregation outcomes that do not correspond with those using naturally-occurring boundaries or randomly-assigned geographies.

In order to create an exogenously-defined geographic unit that includes sparsely-populated areas, we divide the whole country into squares of width and height equal to 0.005 degrees in longitude and latitude degrees respectively. At the average latitude, squares of 0.005 degrees have an approximate width of 555 meters. For 2008 metro areas, the average population of the 28,521 generated grids that we employ in our analysis is 1,074 (s.d.=2,127).[9]

Table 3 presents the summary statistics for the 28,521 neighborhoods in metro areas that

[8]Results are robust to using alternative impedance functions with distance exponents of 0.5, 1, 3 and 4. Larger exponents tend to overweight the relative importance of the closest POIs, whereas smaller ones emphasize access to a numerous portfolio of POIs within a broader catchment area.

[9]This must be contrasted with the average population of 1,452 (s.d.=575) for census tracts. While the physical area of our neighborhoods is not exactly constant across latitudes, it is very close to being so. Note that conventionally-used census tracts are substantially more irregular in shape than our random boundaries, in addition to their endogeneity. Using constant latitude and longitude boundaries makes the neighborhood assignment process for our more than 7 million unique addresses computationally feasible. Note further that metropolitan area fixed effects take care of the very small differences in actual areas between different cities at different latitudes.

will be analyzed in the next section.[10]

(Table 3)

To realize the importance of new neighborhoods, note that 7 percent of the squares in the country's metro grid were empty in 2001 but had become inhabited in 2008 by 234,000 natives and 45,000 immigrants (almost half of them in metro Madrid only). These sprouting neighborhoods were an extreme consequence of the construction boom. In addition, an extra 19 percent of our neighborhoods more than doubled their population between 2001 and 2008. Almost 6 percent of the population in 2008 (1.5 million people) was living in such booming neighborhoods.

Table 3 organizes our control variables by their main source and year of measurement. When possible, variables are measured at baseline —2001— corresponding to that census year. However, our POI gravities (Table 3) are measured in 2008, as detailed data from navigation systems were relatively rare earlier (ESRI, 2009).

3.2 A Statistical Outflow-inflow Equation of Immigrants and Natives

Consider a number of neighborhoods (denoted by subscript k), in a metropolitan area (denoted by m), at a time period denoted by t, and the following naive single equation model:

$$\Delta nat_{k,m,t} = \theta_{m,t} + \beta \triangle im_{k,m,t} + X'_{k,m,t-1}\mathit{\Gamma} + \gamma_n S_{k,m,t} + \varepsilon_{k,m,t} \qquad (1)$$

Changes in the native population between t and $t-1$ here are a function of: a general metropolitan shifter ($\theta_{m,t}$); the change in the number of immigrant arrivals in the neighborhood ($\triangle im_{k,m,t}$); initial neighborhood characteristics the valuation of which may be changing ($X_{k,m,t-1}$); an additional shock to the relative attractiveness of the neighborhood due to changes in employment or amenities ($S_{k,m,t}$); and an i.i.d. random component ($\varepsilon_{k,m,t}$). Because we are using absolute levels in the changes of the dependent variable, note that $X_{k,m,t-1}$ should include controls for initial population scale.[11]

[10] The Appendix details how we get at 28,521 final neighborhoods for the 83 Spanish metro areas identified in Section 2.

[11] It is straightforward to sketch the microfoundations of the demand-side aspects in equation (1) as arising from the addition of outcomes of a simple individual random utility maximization model (RUM) à la McFadden (1974) or from a RUM model with more general error structures, such as the one introduced

The parameter of interest in equation (1) is β, which corresponds to the effect on net native population loss or gain caused by the arrival of immigrants into the neighborhood. We hereinafter denominate this as the average *outflow-inflow* parameter. β is a reduced-form parameter from a general equilibrium process arising from the residential tastes of the different ethnic groups. Because there is a large number of neighborhoods within each metropolitan area, housing prices evolve as compensating differentials across them. Hedonic models (Rosen, 1974) capture this process, and sometimes are specified to include ethnic composition as a local attribute (Wong, 2013). Changes in local ethnic composition may repel (attract) natives. But the subsequent decline (increase) in housing prices and growth (reduction) in home availability generally mutes the full impact of preferences on native demographic change. There is a large extant number of residential general equilibrium models that combine ethnic preferences with housing supply and mobility outcomes. We cannot add much value here to that extensive literature, so we refer the reader to existing segregation models.[12]

Empirically, we focus on the outflow-inflow parameter (β) for two reasons: (i) there is no detailed data for housing prices at the neighborhood level in our context; (ii) more importantly, reduced-form parameters such as β (or their nonlinear counterparts in the tipping-point literature) are of first-order importance for forecasting, policy, and for their social implications. The empirical size of β determines the patterns of spatial segregation and establishes future social and human capital externalities between the native and immigrant populations (Borjas, 1995; Benabou, 1993). Here, we are genuinely interested in the outflow-inflow relationship between immigrant arrivals and native removals from specific neighborhoods, and on what neighborhood characteristics mediate such phenomenon.

There are, however, further interesting implications about the nature of native preferences arising from the empirical sign and magnitude of the ethnic outflow-inflow parameter (β). In a spatial equilibrium, changes in housing supply or residential densities must accommodate new population inflows into a neighborhood in a way such that prices there are equal to

by Bertoli, Brücker, and Fernández-Huertas Moraga (2016)

[12]The literature in tipping points and segregation is rich, starting with Schelling (1971); more recent examples are Bond and Coulson (1989) or Pancs and Vriend (2007). The urban economics literature has also a strong tradition of using segregation models with spatial features, as in Bailey (1959), Yinger (1976), Courant and Yinger (1977), or Kanemoto (1980). The online appendix in Saiz and Wachter (2011) provides a stylized model that incorporates most of the salient issues at arguably low expositional cost.

those in other neighborhoods with identical amenities and accessibility.

If $\beta > 0$, after *controlling for all other relevant neighborhood characteristics and shocks*, one can conclude that the arrival of new immigrants into a neighborhood made it more desirable to the marginal natives moving in. The increased population levels must be accommodated via a combination of new housing supply and growth in local residential densities, with housing prices increasing due to an amenity premium.

With $0 > \beta \geq -1$, immigrant destinations are growing, but the evidence can be consistent with a number of alternative interpretations: (i) a simple mechanical-displacement relationship arises because of tightness in the local housing market; as native families exit randomly, they are replaced by immigrant households living at higher residential densities, while marginal native preferences for the neighborhood have not changed; (ii) a price-displacement story, where immigrant arrivals push up housing prices thereby crowding out some natives, while preserving a population of native marginal residents with higher willingness-to-pay;[13] (iii) a native-flight story, whereby natives dislike the presence of foreigners at the margin, but these preferences are not very strong and lower prices compensate for many of them to stay in the community.

Hypothesis (i) implies housing prices not to be increasing or decreasing in the neighborhoods that receive immigrant inflows and, therefore, new housing construction not to be significantly different from that in identical neighborhoods not receiving immigrants. Hypothesis (ii) implies higher prices and —on average— construction in neighborhoods that see immigrant arrivals. Hypothesis (iii) implies lower prices and —on average— less construction than in identical "control" neighborhoods, with population growth being solely accommodated via increased residential densities of immigrants.

When $\beta < -1$ the neighborhood is losing population and the evidence unambiguously points to the existence of nativist ethnic preferences: the area became less attractive to natives compared to otherwise identical locations, despite the fact that housing there should be easier to get by. Housing prices should be growing more slowly than in comparable neighborhoods "untreated" by an immigration shock.[14]

[13] For instance, the natives staying in the neighborhood may display larger moving costs. Incidentally, this hypothesis also implies higher willingness-to-pay by immigrants for homes in the destination neighborhoods, perhaps because of a premium for ethnic networks.

[14] An alternative explanation for $\beta < -1$ is that immigrants in destination areas live at lower residential densities (less individuals per house), while paying the same or higher prices per square meter. This hy-

3.3 Simultaneity and Omitted Variables

Estimating the model in equation (1) directly is bound to produce biased estimates. Consider the additional equation behind the data generation process for the relative growth of the immigrant population in neighborhood k, of city m, at time t:

$$\begin{aligned}\Delta im_{k,m,t} = \pi_{m,t} + \delta im_{k,m,t-1} + \rho \cdot F(\Delta IM_{m,t}) \cdot im_{k,m,t-1} \\ + X'_{k,m,t-1}\Omega + \gamma_i S_{k,m,t} + \alpha \Delta nat_{k,m,t} + \xi_{k,m,t}\end{aligned} \tag{2}$$

Here, $F(\Delta IM_{m,t})$ stands for a function of the total number of immigrants arrived in the metropolitan area m between t and $t-1$. Immigrants are likely —as natives— to be attracted by local shocks in employment opportunities and amenities ($S_{k,m,t}$), which is likely to bias the estimates of β in equation (1) upwards, having us understate the degree of ethnic displacement. In addition, immigrant inflows are likely to be endogenous to native exits via mechanically taking advantage of new vacancies or reduced competition in the housing market (as captured by the parameter α), which would tend to exaggerate estimates of native flight. The relative strength of these two effects determines the bias of OLS estimates of β in equation 1.

New immigrants are also likely to be attracted to neighborhoods with the presence of co-ethnics. In equation (2), we capture this effect by including the share of the foreign-born at $t-1$ ($im_{k,m,t-1}$). Of course, we should expect more inflows into ethnic neighborhoods in cities that are becoming larger immigrant destinations. This expectation has been empirically corroborated by Saiz and Wachter (2011), and confirmed for the Hispanic population in the US by Kasy (2015): ethnic enclaves expand faster in metropolitan areas where the overall numbers of the relevant minority groups are growing more heavily.

Equation (2) hints to the potential use of temporally lagged immigrant concentrations as instruments for immigrant arrival shocks. Such variables are less likely to impact directly future changes in the native population. In the initial spatial equilibrium, marginal natives should be indifferent between neighborhoods due to compensating differentials, and all advantages/disadvantages of the ethnic composition of a neighborhood should have been

pothesis is counterfactual in the Spanish case, where all the groups with growing presence in the country actually have been shown to live at higher residential densities than comparable native households. We find this using our own data in section 4.3, but this pattern has been amply documented earlier (Onrubia, 2010).

arbitraged away. In other words, if the number of immigrants mattered for native location, the native-flight effect could have taken place already at $t-1$. Furthermore, in the Spanish case, *initial immigrant concentrations were very small and unlikely to have directly changed future perceptions of the neighborhoods by natives.* In contrast, a snowball process with immigrants locating in proximity to friends and family in sequential waves is likely to have led to big disparities in the final level of immigrants across neighborhoods.

Nevertheless, a much less restrictive identification assumption can be obtained by using the interaction between lagged immigrant shares and the metropolitan level of foreign migration $\frac{IM_{m,t}}{IM_{m,t-1}} \times im_{k,m,t-1}$ as an instrument. Here, $IM_{m,t}$ is the total number of immigrants in metro area m at time t. Both lagged minority density and metropolitan fixed effects can then be controlled for directly in a new empirical IV model of native mobility:

$$\Delta nat_{k,m,t} = \theta_{m,t} + \beta \triangle im_{k,m,t} + \rho \cdot im_{k,m,t-1} + X'_{k,m,t-1}\Gamma + \gamma_n S_{k,m,t} + \varepsilon_{k,m,t} \tag{3}$$

We effectively operationalize the estimation of (3) by using "predicted" neighborhood immigrant shocks as instrumental variables. These constructed IVs are derived from original local immigrant densities interacted with migrant growth by country of origin and metro area. Conceptually, the identification relies on comparing neighborhoods that had identical initial levels of ethnic diversity, but that were exposed to different migratory shocks *ex post* by virtue of being in cities that received different overall inflows of the relevant immigrant groups. This is, of course, reminiscent of the Bartik (1991) "shift-share" strategy, this time using differences in aggregate immigrant inflows across metro areas and ethnicities (Saiz and Wachter, 2011; Kasy, 2015). By including lagged immigrant levels $im_{k,m,t-1}$ on the right-hand-side, we directly control for potential delayed demographic responses to initial immigrant settlement patterns, and for unobservables that could be correlated with them. Since we can also control for a very large number of amenities, these types of neighborhoods must have been extremely similar ex ante. Because, in this historical natural experiment, citizens in year 2000 could not have possibly forecasted the large size and speed of future migratory inflows in the country (and less so their ethnic and metropolitan distribution), changing dynamics in these neighborhoods are bound to be generated by unexpected ex post differences in the levels of new immigrant arrivals, having controlled by pre-existing characteristics.

3.4 New Neighborhoods and Segregation Models

Due to the housing boom, many neighborhoods that were empty in 2001 had become heavily populated by 2008. New residential developments were also destinations for the foreign-born. And it was precisely the metropolitan areas receiving large migratory flows that saw more construction (González and Ortega, 2013). It is critical to separately study the demographic dynamics in these new or rapidly-growing neighborhoods.

Generally, sidestepping the analysis of new housing developments may lead researchers to an incomplete picture of ethnic segregation in metropolitan areas where increased residential mobility of minorities is accompanied by general robust demographic growth.

To illustrate this point, consider a theoretical city with 10 neighborhoods, 3 of which are empty in an initial period ($t = 0$) as displayed in Figure 4. We assume the population of the other neighborhoods to be 100 at $t = 0$. The share of minorities in each neighborhood is signified by the relative size of the blue area therein, and their actual location is portrayed at scale. Now consider the pattern of demographic growth at $t = 1$ illustrated in Figure 4. Neighborhood 2, adjacent to a minority enclave "tips" into becoming an enclave itself. However, previously undeveloped neighborhoods (8, 9, 10) are now recipients of inflows of both ethnic groups, and substantial mixing happens there at $t = 1$.

Figure (4)

Empirical tipping point models (Card, Mas, and Rothstein, 2008) study the ability of nonlinearities in past minority shares to predict subsequent changes in the majority population. In this example, such models could find evidence of tipping behavior, since the neighborhood with substantial lagged minority shares transitioned towards complete segregation. But substantial ethnic mixing was happening in the new developments, which paints a more optimistic picture of the evolution of ethnic integration. In our example, subsequent research (data from $t = 2$) could confirm the stability of the mixed-ethnicity patterns in neighborhoods 8, 9, and 10. However, since the micro data used in these studies stems typically from decennial censuses, which themselves have 3-4 year data processing times, this may imply a more than 13-year lag in the potential identification of new integrated urban patterns. Note that even if the neighborhood ethnic compositions at $t = 1$ are sustained at $t = 2$, researchers could mistakenly attribute them to recent changes in tipping point thresholds, as opposed to what they were: well established integrated dynamics in newly-built housing developments.

The same caveat applies to outflow-inflow studies such as Saiz and Wachter (2011). Of course, instruments based on lagged demographic characteristics cannot be obtained (or may be too weak) in areas with no (or very sparse) previous settlement. In this example, using past settlement patterns combined with city-wide growth to form instruments for minority inflows would identify the displacement effect in neighborhood 2, which is very large. However, the estimate would also omit the dynamics in the new residential developments. Existing empirical techniques thus estimate a Local Average Treatment Effect (LATE) that may be relatively less relevant to overall segregation dynamics *in rapidly growing cities*. In our example, the LATE identified by both tipping and outflow-inflow models corresponds to the impact of growing minority concentrations on white flight *in the dense urban areas that have historically bordered minority enclaves*. The types of minority households who move into these areas, and the expectations of non-minority neighbors about the desirability and feasibility of integrated outcomes could be different in older established communities, compared to ex-novo mixed developments.[15] In our work, we will therefore combine the study of the LATE of immigration in densely-settled areas, per equation 3, with descriptive evidence of patterns in rapidly growing suburban developments.

4 Results

4.1 Outflow-inflow: OLS Specification

We estimate an augmented OLS version of equation (1). Empirically, we separate neighborhood inflows of foreign-born individuals from Western Europe and high-income countries (denoted by $\Delta im^{W}_{k,m,t}$) from those from other countries ($\Delta im^{NW}_{k,m,t}$). Our focus is on the later, mainly immigrants from Latin America, Africa, Eastern Europe, and Asia. Foreign-born individuals moving from Western countries (many of them retirees or skilled professionals moving in for lifestyle reasons) are likely to display very different residential dynamics, and never featured prominently on the local debates about the impact of immigration.

Our randomly-defined neighborhoods (squares in a grid) contain many instances of zero or very small population, which dissuades us from normalizing the left-hand variable by dividing net native flows by the local initial population. Alternatively, we control for a

[15] For instance, as in global games, it could be harder for ethnic groups to coordinate into a segregated equilibrium in new areas where clear signals about others' beliefs are forbidden by law.

flexible function (quartic polynomial) of population at baseline ($pop_{k,m,t-1}$), and include a dummy variable for neighborhoods that were empty initially ($empty_{k,m,t-1}$). Our results using this scaling approach are very similar to those using changes in the native population as a share of lagged total population *in the subsample of neighborhoods with a substantial denominator*. The empirical model that we estimate is thus:

$$\Delta nat_{k,m,t} = \theta_{m,t} + \beta^{NW} \Delta im^{NW}_{k,m,t} + \beta^{W} \Delta im^{W}_{k,m,t} + \sum_{i=1}^{4} \kappa_i pop^{i}_{k,m,t-1}$$
$$+\lambda \cdot empty_{k,m,t-1} + A'_{k,m,t}\gamma^{A}_{n} + X'_{k,m,t-1}\Omega^{S}_{n} + u_{k,m,t} \quad (4)$$

where $A_{k,m,t}$ is a vector of amenities (gravities) and $X_{k,m,t-1}$ is a vector of socioeconomic characteristics at baseline. In our application, we have $t = 2008$ and $t-1 = 2001$ and are thus estimating a long-differences model of changes between 2001 and 2008. We believe the low-frequency variance in the data to contain much less error than annual changes. Moreover, local demographic trends were very persistent, and most of the true signal contained in annual changes is well-captured by the long-difference model.

Our population variables ($\Delta nat_{k,m,t}$, $\Delta im^{NW}_{k,m,t}$, $\Delta im^{W}_{k,m,t}$, and $pop_{k,m,t-1}$) are computed excluding children 0-15 years old, in order to avoid capturing native growth due to locally-born children of immigrants. We include mortality and age structure controls: the baseline share of native population in age groups 15-24, 25-44, 45-64, and beyond. We also include controls for the baseline share of immigrants.

Because we have a very large amount of observable neighborhood characteristics, we saturate the model to avoid many of the omitted variable concerns that may plague this kind of study. We will thus control for a dummy for zero population in 2008; distances to population-weighted metro area and municipality centers; and 62 measures of accessibility to amenities, as captured by our gravities.[16] Additionally, we include other variables from the 2001 Census capturing social dimensions and physical attributes that can be differentially relevant for immigrants and natives: unemployment rate (the best indicator of SES in Spain), share of workers in construction, share of housekeeping workers, share of hotel and restaurant workers, share of housing units by building year, average height of the build-

[16]While we control for variables that are measured in 2008 and are therefore potentially affected by the treatment (e.g. POIs), results do not change substantially if we control only for pre-determined variables as measured in 2001.

ings, perceived neighborhood cleanliness, and a number of principal component indices for transportation choices (car, public transportation, walking to work, etc). Table 3 displays all control variables that we use in the study.

(Table 4)

Column 1 of Table 4 shows baseline descriptive results including metro area fixed effects, but no any other controls. Results display a "naive" outflow-inflow coefficient of about -.5 for arrivals from non-Western countries. As we had expected, there was a positive association between the inflows of high-income countries and natives. In column 2, we introduce the control variables, which substantially mute the displacement impact of the immigrants from non-Western countries (this is the coefficient we focus on hereinafter). This finding suggests that observable characteristics were making these neighborhoods counterfactually less attractive to natives, regardless of immigration. Of course, ceteris paribus, unobserved local positive shocks in housing and labor markets could still bias conditional coefficients toward zero in column 2.

To avoid results being driven by economic shock outliers, we next proceed by excluding grid squares with extreme demographic growth or decline in subsequent regressions of Table 4. To define such, we calculate the population growth distribution at the top and bottom 1 percent of the *2001-population-weighted* grid squares. The set of neighborhoods between the top and bottom percentiles experienced population growth between -25.3 percent and 134.8 percent. We now thus exclude 4,531 grid squares with population growth below and above that range. We also drop 1,949 squares with no population in 2001, technically experiencing infinite or indeterminate population growth. The percentage of areas excluded amounts to 22.72 percent of the squares in our metropolitan grids, but to only 2 percent of the metro population in 2001 *by definition*. Remarkably, these neighborhoods went on to encompass 4.9 percent of the metro population in 2008 due to the construction boom, having attracted 631,277 natives and 156,417 immigrants from developing countries. Considering only the grid squares with no such extreme growth or decline (this is, *the winsorized sample* encompassing 98 percent of the country's initial metro population) increases displacement effects to -.27 (column 3). This change suggests that large positive unobserved shocks attracting both groups (e.g. new urban growth and real estate development areas) tended to bias downward outflow-inflow estimates.

Tabulations of the data from the winsorized sample of 22,041 neighborhoods in column

3 display that they received 2.4 million adult immigrants, while 300,000 natives left, in net terms. While we cannot develop causal inference from OLS results, they are useful to establish counterfactual accounting identities.[17] Because the outflow-inflow coefficient in column 3 implies a displacement of about 550,000 natives from their original neighborhoods, taking the results at face value would imply that about 250,000 natives relocated within the winsorized sample; the other 300,000 natives, potentially displaced from established neighborhoods by immigrant arrivals, contributed to 50 percent of the native growth in new real estate developments, which clearly played a key role in this process and will be studied later.

Also using the winsorized sample, Column 4 (Table 4) presents OLS associations by the main regions of origin, focusing on non-Western countries. There was *prima facie* evidence of displacement effects for African, Latin American, and Eastern Europeans. The two latter are precisely the groups that grew more substantially during the period. Nevertheless, standard errors by ethnic group are large and we cannot reject negative displacement effects for all groups.[18]

Column 5 adds the initial share of immigrants from non-Western and Western countries in 2001 as controls, in order to account for unobservables that could be correlated with initial ethnic patterns of settlement. Column 6 uses neighborhood population weights (average population across the initial and final periods). Results are relatively unchanged suggesting a modest displacement effect of one native leaving for each three immigrants from developing countries entering a conventional neighborhood.

4.2 Outflow-inflow: IV Estimation

As exposited earlier, our instrumental variable (IV) strategy builds on recent research (Saiz and Wachter, 2011; Kasy, 2015)), combining spatial diffusion models of existing local settlement patterns of minorities with the shift-share approach in Card (2001), itself inspired by Bartik (1991). Back to equation 2, we use the percentage growth of the immigrant population in each metro area by ethnic group to create our instrument. Concretely, focusing

[17] By counterfactual here we mean that the numbers obtained from estimated outflow-inflow micro relationships are conditional on the large number of observable neighborhood characteristics.

[18] Our instrumental variables approach produces relatively weak instruments for some of the ethnic groups, reducing overall F-stats to below conventional thresholds, and so we focus on the Average Treatment Effect (ATE) of non-Western-country migratory inflows as experienced in this historical event.

on immigrants from non-Western countries we create predictions of total immigrant inflows into neighborhood k, in metro area m, in period t ($\widetilde{\Delta im}^{NW}_{k,m,t}$) as

$$\widetilde{\Delta im}^{NW}_{k,m,t} = \sum_{\forall g} \left(im^g_{k,m,t-1} \cdot \frac{IM^g_{m,t}}{IM^g_{m,t-1}} \right) \tag{5}$$

g denotes an ethnic group, proxied by country of birth, in the subset NW, and $IM^g_{m,t} = \sum_{k \in K^m} im^g_{k,m,t}$, with K^m representing the set of neighborhoods in metro area m. We are therefore projecting the growth rate of each ethnic group in their metro area to forecast its local expansion. Similar calculations are produced to generate instruments for immigration from Western countries. Importantly, our IV specifications will also control for metropolitan area fixed effects, and for the initial concentration of immigrants in each subset (this is $\sum_{\forall g} im^g_{k,m,t-1}$), and are thus not identified by general metropolitan shocks or by initial migrant concentrations per se. Rather, identification relies on *the interaction* between country-by-metro-specific migrant shocks and lagged micro settlement patterns by country and neighborhood.

For instance, consider two hypothetical cities: A and B. Both contain observationally-equivalent neighborhoods 1 and 2. Neighborhood 1 in each city contains a substantial and identical number of Ecuadorian immigrants at $t-1$, while neighborhood two houses a similar contingent of Bolivians. If city A receives larger subsequent inflows of Ecuadorians and city B larger contingents of Bolivians, we would expect neighborhood 1 in city A to be more substantially treated by a higher immigrant dosage, whereas the high-treatment neighborhood in city B would be 2. Note that in a hypothetical regression on the impact of migration on local outcomes we can control for generic metropolitan fixed effects and *for the initial concentrations of immigrants in each neighborhood*. Results from this shift share IV 2SLS specifications are presented in Table 5.

(Table 5)

Column 1 and column 2 present identical specifications using the full (unwinsorized) sample of grid squares, with the latter column using average 2001-2008 population weights. Our causal estimates suggest that every 3 immigrants from non-Western countries arriving into a well-established urban neighborhood displaced one native. There is some evidence of complementarity with regards to Western arrivals, but the coefficient is much smaller and not always significant, consistent with the existence of local shocks attracting both kinds of populations in the OLS specification.

In columns 3 and 4, we introduce IV results for two separate neighborhood subsamples with unemployment rates below and above the median in the 2001 Census, respectively. In stark contrast with results in the US context (Saiz and Wachter, 2011), we do not find evidence of stronger "native flight" in the neighborhoods with higher socio-economic status. In fact, displacement effects seem slightly larger in the neighborhoods with higher initial unemployment, but differences are not statistically significant.

As we discussed in section 3.4, the results represent the LATE of an exogenous immigrant shock into neighborhoods of traditional settlement (as opposed to new developments). While slightly higher than OLS estimates, the outflow-inflow parameter is substantially below one and larger than zero in absolute terms. The results are thus consistent with an absence of strong nativist residential behavior *at the margin*. But they are also consistent with mild native flight that was quickly compensated by lower housing price growth, with all the adjustment in quantities coming from higher residential densities of immigrant households. This hypothesis would imply relatively less new housing supply in treated (high-immigration) neighborhoods, compared to neighborhoods without immigrants, where prices should have been growing faster.

4.3 New Construction in Settled Neighborhoods

While we do not have data on housing prices at the micro level during this period, we do have a measurement of construction activity: all new street addresses had to be duly registered by the municipal government and appear in our dataset as soon as the first residents move in. Thus, we do not have a measure of the number of new homes built, but we know how many buildings (corresponding to a unique street address) were erected or removed from the stock. As implied by conventional urban models, the number of homes in multifamily units tends to be larger in central locations and smaller in far away locales, with single-family units becoming more common in the suburbs. Therefore a new building (signified by a new street address) does not correspond to the same number of homes across geographies. Yet we can make inferences based on the initial inhabitants per building by neighborhood.

Concretely, we can express the number of homes in the neighborhood in 2001 as a function of the population and a location-specific parameter $\delta_{k,m}$ capturing the inverse of the average number of persons per home in the grid square: $Homes_{k,m,t-1} \equiv Pop_{k,m,t-1} \cdot \delta_{k,m}$. Similarly, the number of initial homes in the neighborhood can be expressed as the number of street addresses times the number of homes per address in the neighborhood ($\tau_{k,m}$): $Homes_{k,m,t-1} \equiv Buildings_{k,m,t-1} \cdot \tau_{k,m}$. Therefore, the initial number of buildings in each neighborhood can be expressed:

$$Buildings_{k,m,t-1} \equiv \frac{Pop_{k,m,t-1}}{(\tau_{k,m}/\delta_{k,m})} \tag{6}$$

Let us consider the hypothesis that potential native population growth happens at established densities, but that non-Western immigrant households move in at higher ones. More precisely, denote the ratio of homes per immigrant person by $\psi\delta_{k,m}$, with $1 > \psi > 0$. We then expect the number of Buildings in our database to grow according to:

$$\Delta Buildings_{k,m,t} = \frac{\Delta nat_{k,m,t}}{(\tau_{k,m}/\delta_{k,m})} + \frac{\psi \Delta im^{NW}_{k,m,t}}{(\tau_{k,m}/\delta_{k,m})} \tag{7}$$

We are still modeling native outflows as a function of immigrant inflows, and other factors ($\nu_{k,m,t}$) as:

$$\Delta nat_{k,m,t} = \beta \Delta im^{NW}_{k,m,t} + \nu_{k,m,t} \tag{8}$$

Combining (7) with (8) and (6) yields:

$$\Delta Buildings_{k,m,t} = (\psi + \beta) \frac{\Delta im^{NW}_{k,m,t}}{(Pop_{k,m,t-1}/Buildings_{k,m,t-1})} + \widetilde{\nu_{k,m,t}} \tag{9}$$

Back to Table (5), this suggests using a similar IV specification as in column 2, in order to model the net change in the number of buildings (new construction minus demolitions) as a function of immigrant inflows divided by original population densities by building in the neighborhood. The instruments —predicted immigrant inflows— are similarly adjusted.

Results are shown in column 5. They suggest that areas receiving a number of immigrants that would have required one extra building at previous native settlement densities, actually saw the appearance of 0.4 new buildings. Because around -0.32 buildings had been vacated by natives, this implies that $\psi = 0.72$. Therefore, a way to rationalize the results is with immigrant residential densities that were 39 percent higher than those of native households.

According to Ajenjo, Blanes, Bosch, Parella, Recio, Martin, and Sintes (2008), each non-Western immigrant housing unit tended to be occupied by 3.85 people, compared to 2.8 people-per units for natives: this corresponds to densities that are 37 percent higher in foreign-born households. Similarly, Reher, Alcala, Quinones, Requena, Dominguez, Gimeno, and Stanek (2008), report the average size of the immigrant household at 3.8 persons compared to 2.9 for natives, and the average size of the corresponding housing units as 96 sq. m. and 97.1 sq. m. respectively, implying a 31 percent higher residential density for immigrants. Both magnitudes are close to our estimates, implying that a local combination of: (i) mild native substitution, (ii) new housing constructions, and (iii) increased residential densities, explain almost perfectly the accommodation of immigrants in the housing market. Note that, per our discussion in section 3.2, these facts are highly inconsistent with a model of generalized nativism, where we would expect less construction due to reduced demand by the marginal native and shrinking relative prices in immigrant destinations.

4.4 Nonlinear Effects

The previous specifications assume a linear relationship between immigrant inflows and native outflows. We now investigate potential non-linear effects. We first search for potential tipping points following the methods in Card, Mas, and Rothstein (2008). We start by working on our winsorized sample[19] and dividing the set of neighborhoods into two equally-sized random samples: training and replication. In the training sample we run a sequence of regressions of the form:

$$\widehat{\Delta nat_{k,m,t}} = \theta'_{m,t} + d \cdot \mathbf{1}\left[shim^{NW}_{k,m,t-1} > \mu^*\right] + \widehat{\varepsilon}_{k,m,t} \quad (10)$$

for $0 \leq \mu^* \leq M$. Here $shim^{NW}_{k,m,t-1}$ denotes the population share of non-Western immigrants in 2001 and $\widehat{\Delta nat_{k,m,t}}$ the covariate-adjusted change in the native population in each neighborhood. This variable is adjusted by estimating regressions with the same covariates as in table 4, obtaining the residuals, and adding those to the average value. Thus we "partial out" the influence of other covariates on $\widehat{\Delta nat_{k,m,t}}$, including scaling to polynomials in initial population and physical characteristics of the neighborhoods. μ^* is the level of initial migrant share that could propiciate a discontinuity in native population growth, and $\mathbf{1}[\cdot]$ denotes an indicator variable. We set $M = 0.50$ and experiment with values of μ^* for all integers between 1 and 49 percent. We then select the μ^* with maximal R-square from the series of training regressions, which turns out to be four percent.

Once this candidate for a discontinuity point is identified we test its relevance on the replication sample. To ensure that the roughly-identified candidate for a "jump" in the de-

[19] Outlier observations have typically zero or very low population base and we cannot credibly identify differences in the initial share of immigrants.

pendent variable represents a true discontinuity we now also control for a quadratic function of the initial immigrant share on "both sides" of $\mu^* = 0.04$ ($f^{+,-}_{0.04}\left(shim^{NW}_{k,m,t-1}\right)$) as shown below:

$$\widehat{\Delta nat_{k,m,t}} = \theta"_{m,t} + \Phi{\cdot}\mathbf{1}\left[shim^{NW}_{k,m,t-1} > 0.04\right] + f^{+,-}_{0.04}\left(shim^{NW}_{k,m,t-1}\right) + \widehat{\zeta}_{k,m,t} \qquad (11)$$

Our out-of-sample estimate of Φ rejects a discontinuity in native growth at conventional significance levels: the estimate yields a positive 6.9 with and estimated standard error of 29.83.

(Figure 5)

The lack of major tipping dynamics with regards to initial immigrant concentrations in 2001 can be appreciated graphically in figure 5. On the horizontal axis, we group neighborhoods into 50 bins defined by their initial share of immigrants in 2001, rounded to the closest integer. In the vertical axis, the dots show average change in the covariate-adjusted number of natives within each bin across neighborhoods. The continuous line fits a local-mean smoothed plot of the relationship using the full dataset, with 95 percent confidence intervals displayed throughout the gray area. Contrary to conventional tipping dynamics, we see a u-shaped relationship: the number of natives seemed to decrease with growing initial immigrant concentrations up to about 10 percent, but then increased for even larger initial immigrant densities.[20]

The lack of clear tipping dynamics may be due to low power, since just a few neighborhoods had significant concentrations of non-Western migrants in 2001: for instance only 5 percent had migrant concentrations above 9.5 percent, and the top percentile was at around a 20 percent migrant share. Although we have more than 20,000 neighborhoods, the phenomenon may not have been well-established enough to display discernible tipping patterns. In contrast, subsequent new arrivals were substantial enough to push many areas beyond potential tipping points. For instance, the neighborhood at the 75 percentile in 2008 already had more than a 10 percent immigrant share.

Of course, the final immigrant share in 2008 was directly endogenous to native flows

[20]We use the Epanechnikov kernel and the ROT method to estimate the optimal bandwidth. The U shape becomes more pronounced, with similar inflection points, if we use unadjusted changes in native populations on the y axis. This later finding confirms the regression results that areas with initial concentrations of immigrant between 0-10 percent tended to have characteristics that would have predicted native decline regardless of migratory inflows.

between 2001 and 2008. In order to find potential exogenous discontinuity points in native residential responses we redeploy our instrumental variables in the following way: we again divide the data into two random search and validation subsamples, expected to be of equal size. We then generate a sequence of tipping-point indicator variables to use as potential instruments, taking the form:

$$\widetilde{TT_{k,m,t}^{NW}(\mu^{**})} = \mathbf{1}\left[\frac{im_{k,m,t-1}^{NW} + \widetilde{\Delta im_{k,m,t}^{NW}}}{pop_{k,m,t-1}} > \mu^{**}\right] \tag{12}$$

Where $\widetilde{\Delta im_{k,m,t}^{NW}}$ is the shift-share instrument produced in equation 5. We thus examine what happened to neighborhoods that, in 2001, had specialized in the nationalities which subsequently grew in their metropolitan area, and whose immigrant shares were thus expected to mechanically exceed μ^{**} by 2008.[21] To do so we run a series of 2SLS specifications on the search sample as in column 2 in Table 5 but adding as an additional control an indicator function for neighborhoods that had actually reached the potential tipping threshold μ^{**} as of 2008 ($TT_{k,m,t}^{NW}(\mu^{**}) = \mathbf{1}\left[\frac{im_{k,m,t-1}^{NW}+\Delta im_{k,m,t}^{NW}}{pop_{k,m,t-1}} > \mu^{**}\right]$) instrumented by $\widetilde{TT_{k,m,t}^{NW}(\mu**)}$. We search for jumps at immigrant densities that reach 5 percent of the initial population and a series of thresholds at additional five percentage points up to 45 percent. In this sequence of 2SLS models applied to the search sample, we find that the μ^{**} that maximizes the t-statistic of the parameter on $TT_{k,m,t-1}^{NW}(\mu)$ corresponds to $\mu^{**} = 0.3$ (30 percent immigrant share). Note that the models include now all control variables, making it less likely for nonlinearities in the immigrant share to capture differences in neighborhood characteristics.

Contrary to tipping dynamics, the coefficient in the search sample happens to be positive (this is, there was a potential positive jump in native inflows whenever immigrant levels reached 30 percent). We then re-estimate the 2SLS model in table 5, column 2, this time in the replication sample and adding a dummy capturing neighborhoods with 30 percent immigrant shares in 2008, instrumented using $\widetilde{TT_{k,m,t}^{NW}}(0.3)$. The new instrument is strong. Yet the coefficient on the change of the native population is still a *positive* 9.29, with standard error of 116.37. There is thus no evidence of a large acceleration of native exodus in neighborhoods where immigrant shares exogenously reached relatively large dimensions.

[21] Remember that we did not find signs of tipping dynamics earlier with a reduced number of neighborhoods that had crossed potential tipping points in 2001. Of course, the new specification is reduced to the sample of neighborhoods with some population in 2001, and we therefore exclude newly-settled areas.

4.5 New Suburbs

Results in the previous subsections are consistent with the existence of very mild native displacement, coupled with increased construction and higher immigrant residential densities in established destination neighborhoods. But even these results understate the degree of demographic complementarity between immigrant and native inflows in this historical episode. Concurrently with the massive immigrant inflow, there was a real estate development boom in the country. As illustrated in Figure 4, results of empirical specifications that use past ethnic settlement patterns only reflect the LATE of immigrant arrivals on native outflows in established neighborhoods, which tend to be denser.

In order to have a full picture we need to describe the contemporaneous patterns of native and migrant settlement in new construction areas. As we saw earlier, net native metropolitan growth happened disproportionately in relatively few booming developments. For instance, large cities (above 100,000 inhabitants in 2001) lost around 0.5 million natives while attracting 1.5 million immigrants. In contrast, metropolitan municipalities with less than 100,000 inhabitants (mostly suburbs or satellite cities in major conurbations) gained both 0.8 million natives and 1.1 million immigrants.

In our analysis, we find a very strong spatial dimension to the unbalanced growth of the native population. Figures 6 and 7 plot the neighborhoods gaining natives (top) and, because most neighborhoods have positive immigrant entries, those neighborhoods with *above average* migrant growth (bottom), in the Madrid and Barcelona metro areas respectively.[22]

(Figure 6)

(Figure 7)

In the figures, each circle corresponds to the centroid of each of our grid squares, and its size is proportional to the 2001 population. We color squares in red according to either the gain of natives (upper panel) or to the above-average number of migrant arrivals (lower panel). The upper panels clearly shows that native growth was more widespread far away from the city centers (mostly taking place in newly-built suburban developments). The lower panels, showing above-average migrant growth (more than 780 migrants per grid square in Madrid and 845 in Barcelona), are almost the inverse image.

In Figure 8 we show that these patterns extended to the full sample of metro areas. Here we display smooth plots of average number of immigrant and native net growth by

[22]Maps for other Spanish metro areas available from the authors upon request.

neighborhood, sorted by the distance to their metropolitan area's central city. The patterns are stark. Metro areas saw average declines in the native population up to around 3.5 kilometers from the center, where immigrant arrivals from non-Western countries were at their highest. After that distance, both native and migrant flows were positive, the former increasing monotonically with distance, the latter decreasing. We can thus speak of two types of neighborhoods: central, denser, established neighborhoods, experiencing mild native flight, and suburban new-development neighborhoods, attracting both migrants and natives.

Native displacement therefore happened more conspicuously in dense areas of the largest cities, in a few instances even generating the appearance of immigrant ghettos. However, the *average* displacement effect, even in central cities, was not inordinately large. In addition, substantial ethnic mixing was happening in less conspicuous locations in the suburbs or in satellite cities.

5 Discussion and Conclusions

The settlement of immigrants into urban areas of the developed world has become a central policy issue of our times. In many countries, a substantial fraction of citizens are expressing dissatisfaction with the growing number of foreign-born neighbors by voting for policy alternatives geared to curtailing such inflows. Other citizens may be voting with their feet by disproportionately fleeing the neighborhoods where immigrants move into. In this context, it is important to understand well the residential dynamics associated with one of the largest and fastest immigration episodes in recent history. International migration in Spain increased from 3 to 13 percentage points of the population between 1998 and 2008, offering a unique opportunity to examine how native residential location decisions are taken, and the impact of immigration on urban form.

We provide a causal estimate of what we define as the outflow-inflow parameter linking net native population decline in a neighborhood to immigrant arrivals from developing countries. This reduced-form parameter is critical to understanding and forecasting residential segregation patterns. It is also an element to inform expectations about the strength of native residential tastes with regards to foreign-born populations. All other local neighborhood attributes and contemporaneous shocks being equal, an outflow-inflow parameter smaller than -1 is a clear indication of native flight and segregationist tastes by at least some natives. In contrast, positive parameters signify complementarity and that some natives prefer immigrant neighbors.

An outflow-inflow parameter of -1 is consistent with no segregationist behavior and a simple mechanical displacement: the space left behind by a native leaving randomly is occupied by an immigrant. While this situation may not imply native animus, it could lead to stark segregation wherever immigrants tend to cluster spatially. A outflow-inflow causal parameter of -1 is also consistent with some native discrimination that is immediately arbitraged away by lower housing prices in the neighborhood, implying relatively less new construction there. Finally, negative parameters between 0 and -1 are also both consistent with mild native flight (implying lower prices and construction levels) or with no segregationists tastes combined with housing market competition (and more construction), both situations implying increased residential densities.

The empirical analysis represents a comprehensive and unique study, combining microdata on the universe of every residents' exact address with the distance of each building to a large portfolio of amenities, and with socioeconomic characteristics of the neighborhood. To the best of our knowledge, this represents a more complete set of controls than what has heretofore been used in the immigrant segregation literature. In addition, and to further elicit a causal effect, we deploy a shift-share instrumental variable similar to Card (2001) but with a micro neighborhood focus (Saiz and Wachter, 2011; Kasy, 2015). Our constructed IVs predict the number of new immigrants in each previously-settled neighborhood starting with its number of foreign-born individuals by nationality at baseline, and projecting their growth using demographic inflows by nationality in the metro area.

Our main finding is that for every three immigrants entering into a settled neighborhood, one native moved away from it. Using discretized versions of our instruments, we do not find evidence of nonlinearities in this displacement effect. We also show that areas receiving a number of immigrants that would have required one extra building at previous native settlement densities actually saw the additional construction of 0.4 new buildings, compared to similar neighborhoods. They also imply immigrant residential densities (number of persons per household) that were 40 percent higher than native ones in destination neighborhoods. The latter result is exactly consistent with measurements by other authors using household surveys.

The combination of very mild displacement with substantial new construction in destination neighborhoods is at odds with the existence of strong segregationist behavior, and implies very robust housing markets in the destination areas. Besides their intrinsic social importance, these results underline the relevance of considering new construction and potential differential densities associated with the arrival of minority groups.

The results also point to the fact that conventional outflow-inflow or tipping-point empirical methods measure local average treatment effects in existing settled, denser communities. While such estimates are of considerable importance in the most visible areas of the city, they may understate the general degree of ethnic mixing in growing metros if, as we find here, new residential developments attract both majority and minority populations. Conventional methods based on lagged demographics may therefore provide a complete picture in Cleveland, Chicago, or Marseille, but are less suitable to understanding patterns of segregation in booming Phoenix, Dublin, or Shenzen.

The large population inflows in the historical episode that we consider actually facilitated the suburbanization of the native population. Most autochthonous urban demographic growth happened in new developments of suburban areas. While immigrants were more likely to locate in central cities, they also found their way into the new developments in substantial numbers. Therefore, native and immigrant inflows appeared complementary in the new booming suburbs.

The absence of strong segregationist behavior may or may not imply that ingrained native preferences for ethnic homogeneity were low among the native population. The speed and magnitude of the immigrant inflows may have made it difficult to forecast future immigrant concentrations in alternate neighborhoods within the city, thereby adding uncertainty to native moving decisions. As recently demonstrated by Blair (2014) the *marginal* taste parameters elicited by reduced-form methods in the segregation literature are conditional to the contextual choice set: agents consider the abundance of similar neighborhoods with low minority densities before considering flight from their current homes. If, for instance, alternate available neighborhoods that are accessible to employment opportunities and amenities all have substantial minority shares, it may not pay off for marginal majority individuals to leave their current neighborhood, even in the presence of ethnic preferences [23]

[23] Of course there may be other neighborhoods with better amenities and no minorities available, but these may be infra-marginal in the sense that —even absent ethnic dynamics— minorities and the marginal majority individuals with which the may overlap geographically will be priced out of them regardless (Blair, 2016).

In this regard, changing expectations about the new ethnic-mixing regime in the country could have made it very difficult for natives to flee. As in macroeconomic and financial global games (Carlsson and Van Damme, 1993) or in network games (Jackson and Yariv, 2007), the decision to flee by natives may endogenously depend on the formation of higher order beliefs about what other natives in the neighborhood or network will expect. Native flight may be more likely whenever most natives believe that the neighborhood will tip and the destination neighborhoods are expected to be safe from ethnic mixing.

The extraordinary magnitude of immigrant arrivals makes this case very interesting but quite unique. It may have been very difficult to acquire common knowledge about native beliefs on the equilibrium ethnic composition of thousands of neighborhoods across the countries' urban system. In other environments, with smaller but steady inflows, the previous history of settlement may provide clues to form expectations. For instance, if majority individuals see substantial ethnic encroachment in a neighborhood that is adjacent to a minority ghetto, they may coordinate in a belief that full segregation will eventually be the outcome, and flee.[24] Self-fulfilling beliefs about ghettoization in these scenarios could thus imply displayed behavioral tolerance to be below psychological tolerance: majority individuals may be fleeing at minority concentration levels that do not cause them contemporaneous psychological discomfort, but which anticipate *future* minority prevalence. In equilibrium, these expectations might turn out to be systematically right.

The environment in our research was likely to make native coordination difficult. Actual psychological tolerance could be smaller than the considerable behavioral tolerance that our results reveal. After all, natives did show concern about the growth of the foreign-born population in surveys. Nonetheless, these sentiments have not crystallized in a strong anti-immigrant political movement so far. The processes surrounding the formation of beliefs around neighborhood ethnic change, and their interaction with the speed and past experience of minority inflows in the city represent fascinating topics that future research in this literature should seriously address.

[24]It is even possible that the expectations about full minority ghettoization are driven by previous explicitly segregationist policies.

References

ACCETTURO, A., F. MANARESI, S. MOCETTI, AND E. OLIVIERI (2014): "Don't stand so close to me: The urban impact of immigration," *Regional Science and Urban Economics*, 45(C), 45–56.

AJENJO, M., A. BLANES, J. BOSCH, S. PARELLA, A. RECIO, J. S. MARTIN, AND E. SINTES (2008): *Les condicions de vida de la poblacio immigrada a Catalunya*, Col.leccio Politiques: Fundacio Jaume Boffill. Editorial Mediterranea, Barcelona.

BAILEY, M. J. (1959): "Note on the Economics of Residential Zoning and Urban Renewal," *Land Economics*, 35(3), 288–292.

BALLESTER, C., AND M. VORSATZ (2014): "Random Walk-Based Segregation Measures," *The Review of Economics and Statistics*, 96(3), 383–401.

BARTIK, T. J. (1991): *Who Benefits from State and Local Economic Development Policies?*, no. wbsle in Books from Upjohn Press, W.E. Upjohn Institute for Employment Research. W.E. Upjohn Institute for Employment Research.

BAUM, C. F., M. E. SCHAFFER, AND S. STILLMAN (2014): "IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation," Statistical Software Components, Boston College Department of Economics.

BELL, B., AND S. MACHIN (2013): "Immigrant Enclaves And Crime," *Journal of Regional Science*, 53(1), 118–141.

BENABOU, R. (1993): "Workings of a City: Location, Education, and Production," *The Quarterly Journal of Economics*, 108(3), 619–652.

BERTOLI, S., H. BRÜCKER, AND J. FERNÁNDEZ-HUERTAS MORAGA (2016): "The European crisis and migration to Germany," *Regional Science and Urban Economics*, 60(C), 61–72.

BERTOLI, S., AND J. FERNÁNDEZ-HUERTAS MORAGA (2013): "Multilateral resistance to migration," *Journal of Development Economics*, 102(C), 79–100.

Blair, P. Q. (2014): “The Effect of Outside Options on Neighborhood Tipping Points,” Ph.D. thesis, Ph. D. thesis, University of Pennsylvania.

——— (2016): “The Effect of Outside Options on Neighborhood Tipping Points,” .

Blázquez, M., C. Llano, and J. Moral (2010): “Commuting Times: Is There Any Penalty for Immigrants?,” *Urban Studies*, 47(8), 1663–1686.

Boeri, T., M. De Philippis, E. Patacchini, and M. Pellizzari (2014): “Immigration, Housing Discrimination and Employment,” CReAM Discussion Paper Series 1414, Centre for Research and Analysis of Migration (CReAM).

Bond, E. W., and N. Coulson (1989): “Externalities, filtering, and neighborhood change,” *Journal of Urban Economics*, 26(2), 231 – 249.

Borjas, G. J. (1995): “Ethnicity, Neighborhoods, and Human-Capital Externalities,” *The American Economic Review*, 85(3), 365–390.

——— (2013): “The analytics of the wage effect of immigration,” *IZA Journal of Migration*, 2(1), 22.

Borjas, G. J., and J. Monras (2016): “The Labor Market Consequences of Refugee Supply Shocks,” NBER Working Papers 22656, National Bureau of Economic Research, Inc.

Bosch, M., M. A. Carnero, and L. Farré (2010): “Information and discrimination in the rental housing market: Evidence from a field experiment,” *Regional Science and Urban Economics*, 40(1), 11–19.

Brücker, H., S. Fachin, and A. Venturini (2011): “Do foreigners replace native immigrants? A panel cointegration analysis of internal migration in Italy,” *Economic Modelling*, 28(3), 1078–1089.

Card, D. (2001): “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration,” *Journal of Labor Economics, University of Chicago Press*, 19(1), 22–64.

CARD, D., A. MAS, AND J. ROTHSTEIN (2008): "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 123(1), 177–218.

CARLSSON, H., AND E. VAN DAMME (1993): "Global games and equilibrium selection," *Econometrica: Journal of the Econometric Society*, pp. 989–1018.

CEA D'ANCONNA, M. A., AND M. S. VALLES MARTÍNEZ (2014): *Evolución del racismo, la xenofobia y otras formas conexas de intolerancia en España [Informe 2013]*. Subdirección General de Información Administrativa y Publicaciones. Ministerio de Empleo y Seguridad Social. Madrid.

CLEMENS, M. A., AND J. HUNT (2017): "The Labor Market Effects of Refugee Waves: Reconciling Conflicting Results," NBER Working Papers 23433, National Bureau of Economic Research, Inc.

COURANT, P. N., AND J. YINGER (1977): "On models of racial prejudice and urban residential structure," *Journal of urban Economics*, 4(3), 272–291.

CUTLER, D. M., E. L. GLAESER, AND J. L. VIGDOR (2008): "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation," *The Review of Economics and Statistics*, 90(3), 478–497.

DAMM, A. P. (2009): "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence," *Journal of Labor Economics*, 27(2), 281–314.

DE LA ROCA, J. (2017): "Selection in initial and return migration: Evidence from moves across Spanish cities," *Journal of Urban Economics*, 100, 33–53.

DE LA ROCA, J., AND D. PUGA (2017): "Learning By Working In Big Cities," *Review of Economic Studies*, 84(1), 106–142.

DOCQUIER, F., A. OZDEN, AND G. PERI (2014): "The Labour Market Effects of Immigration and Emigration in OECD Countries," *The Economic Journal*, 124(579), 1106–1145.

DUSTMANN, C., AND I. P. PRESTON (2007): "Racial and Economic Factors in Attitudes to Immigration," *The B.E. Journal of Economic Analysis & Policy*, 7(1), 1–41.

Echazarra, A. (2010): "Segregación residencial de los extranjeros en el área metropolitana de Madrid. Un análisis cuantitativo," *Revista Internacional de Sociología (RIS)*, 68(1), 165–197.

Edin, P.-A., P. Fredriksson, and O. Åslund (2003): "Ethnic Enclaves And The Economic Success Of Immigrants - Evidence From A Natural Experiment," *The Quarterly Journal of Economics*, 118(1), 329–357.

ESRI (2009): *StreetMap Premium Europe NAVTEQ 2009 Release 2*.

Fernández-Huertas Moraga, J., A. Ferrer, and A. Saiz (2011): "Localización de los inmigrantes y preferencias residenciales de la población autóctona: Nuevos guetos?," mimeo.

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

García-López, M. (2012): "Urban spatial structure, suburbanization and transportation in Barcelona," *Journal of Urban Economics*, 72(2-3), 176 – 190.

Gautier, P. A., A. Siegmann, and A. Van Vuuren (2009): "Terrorism and attitudes towards minorities: The effect of the Theo van Gogh murder on house prices in Amsterdam," *Journal of Urban Economics*, 65(2), 113–126.

Giulietti, C. (2009): "Immigration and displacement across local labour markets," Discussion Paper Series In Economics And Econometrics, Economics Division, School of Social Sciences, University of Southampton 0917, Economics Division, School of Social Sciences, University of Southampton.

González, L., and F. Ortega (2013): "Immigration And Housing Booms: Evidence From Spain," *Journal of Regional Science*, 53(1), 37–59.

Hatton, T. J., and M. Tani (2005): "Immigration and Inter-Regional Mobility in the UK, 1982-2000," *Economic Journal*, 115(507), F342–F358.

INE (2005): *Censo de Población y Viviendas. Volumen IV. Indicadores inframunicipales. Secciones censales y unidades poblacionales CD Rom*. Madrid.

——— (2009): *Microdatos del Padrón Municipal de Habitantes*. Madrid.

Jackson, M. O., and L. Yariv (2007): "Diffusion of behavior and equilibrium properties in network games," *The American economic review*, 97(2), 92–98.

Jofre-Monseny, J., M. Dahlberg, and P. Fredriksson (2012): "On the dynamics of segregation," ERSA conference papers ersa12p832, European Regional Science Association.

Kanemoto, Y. (1980): "Externality, migration, and urban crises," *Journal of urban Economics*, 8(2), 150–164.

Kasy, M. (2015): "Identification in a model of sorting with social externalities and the causes of urban segregation," *Journal of Urban Economics*, 85, 16–33.

Massey, D. S., and N. A. Denton (1988): "The Dimensions of Residential Segregation," *Social Forces*, 67(2), 281–315.

Maza, A., J. Villaverde, and M. Hierro (2013): "Explaining the settlement patterns of foreigners in Spain," *Applied Geography*, 40(0), 11 – 20.

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.

Ministerio de Vivienda (2007): *Atlas Estadístico de las Áreas Urbanas de España 2006*. Centro de Publicaciones. Secretaría General Técnica. Ministerio de Vivienda. Madrid.

Nicodemo, C., and J. M. Raya (2012): "Change in the distribution of house prices across Spanish cities," *Regional Science and Urban Economics*, 42(4), 739–748.

Onrubia, J. (2010): "Vivienda e inmigración en España: situación y políticas," *Presupesto y Gasto Público*, (61), 273–310.

Pancs, R., and N. J. Vriend (2007): "Schelling's spatial proximity model of segregation revisited," *Journal of Public Economics*, 91(12), 1 – 24.

Pareja-Eastaway, M. (2009): "The Effects Of The Spanish Housing System On The Settlement Patterns Of Immigrants," *Tijdschrift voor Economische en Sociale Geografie*, 100(4), 519–534.

Reher, D.-S., L. C. Alcala, F. G. Quinones, M. Requena, M. I. S. Dominguez, A. S. Gimeno, and M. Stanek (2008): *Informe Encuesta Nacional de Inmigrantes (ENI - 2007)*, INE Documentos de Trabajo. Instituto Nacional de Estadistica, Madrid.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82(1), 34–55.

Ruist, J., J. Stuhler, and D. A. Jaeger (2017): "Shift-Share Instruments and the Impact of Immigration," .

Saiz, A., and S. Wachter (2011): "Immigration and the Neighborhood," *American Economic Journal: Economic Policy*, 3(2), 169–88.

Schelling, T. C. (1971): "Dynamic models of segregation?," *Journal of mathematical sociology*, 1(2), 143–186.

Verdugo, G. (2011): "Public Housing and Residential Segregation of Immigrants in France, 1968-1999," *Population (english edition), Institut National d'Études Démographiques (INED)*, 66(1), 169–193.

Wong, M. (2013): "Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore," *Review of Economic Studies*, 80(3), 1178–1214.

World Bank (2013): *World Development Indicators*. http://databank.worldbank.org.

Yinger, J. (1976): "Racial prejudice and racial residential segregation in an urban model," *Journal of urban economics*, 3(4), 383–396.

A Tables and Figures

Table 1: Evolution of the Spanish native and foreign-born population in the Municipal Registry

Year	Population	Immigrants	Share
1998	39,852,650	1,173,767	2.9%
1999	40,202,158	1,259,054	3.1%
2000	40,499,790	1,472,458	3.6%
2001	41,116,842	1,969,269	4.8%
2002	41,837,894	2,594,052	6.2%
2003	42,717,064	3,302,440	7.7%
2004	43,197,684	3,693,806	8.6%
2005	44,108,530	4,391,484	10.0%
2006	44,708,964	4,837,622	10.8%
2007	45,200,737	5,249,993	11.6%
2008	46,157,822	6,044,528	13.1%

Source: own elaboration on data from INE (2009)

Table 2: Place of Birth of Spanish immigrants' stocks and flows (2001-2008)

	Share of Total Immigration from:					
Year	**Eastern Europe**	**Latin America**	**Subsaharian Africa**	**Maghreb**	**Rest of Asia**	**Most Developed**
1998	1.9%	24.5%	2.9%	17.7%	5.3%	47.7%
1999	1.9%	24.6%	3.2%	16.5%	6.9%	46.9%
2000	2.7%	26.5%	3.5%	17.4%	4.6%	45.2%
2001	5.2%	32.8%	3.7%	16.7%	4.4%	37.2%
2002	7.5%	37.7%	3.6%	15.9%	4.3%	31.1%
2003	9.9%	40.3%	3.4%	14.7%	4.3%	27.3%
2004	11.9%	42.2%	3.4%	14.2%	4.3%	24.1%
2005	13.9%	40.9%	3.6%	14.0%	4.6%	23.0%
2006	15.1%	39.3%	3.7%	13.7%	4.9%	23.2%
2007	16.9%	38.7%	3.5%	12.9%	4.6%	23.4%
2008	19.2%	38.0%	3.4%	12.3%	4.7%	22.3%
Immigrants in 2008	1,161,290	2,298,787	208,497	745,788	281,925	1,348,241
2001-2008 increase	1,058,802	1,652,565	135,755	416,610	194,503	616,435
Shares	26.0%	40.6%	3.3%	10.2%	4.8%	15.1%

Source: own elaboration on data from INE (2009)

Table 3: Summary Statistics for Neighborhoods in Spanish Metro Areas (2001-2008)

Variable	Average	Std. Dev.	Min.	Max.
$\Delta nat_{k,m,2001\_2008}$	12.99	299.97	-4,517	5,160
$\Delta im^{NW}_{k,m,2001\_2008}$	86.88	239.89	-1,578	5,964
$\Delta im^{NW}_{k,m,2001\_2008}$	4.43	43.27	-1,423	3,061
2001 Variables				
$pop_{k,m,2001}$	821.87	1778.46	0	25,139
NW immigrants in 2001	43.92	126.06	0	3,226
W immigrants in 2001	14.68	49.60	0	2,780
No population in 2001	0.07	0.25	0.00	1.00
Share of pop. Aged 15-24	0.13	0.09	0.00	1.00
Share of pop. Aged 25-44	0.31	0.15	0.00	1.00
Share of pop. Aged 45-64	0.21	0.13	0.00	1.00
Share of pop. Aged 65+	0.14	0.14	0.00	1.00
Unemployment rate	11.56	6.83	0.00	68.33
Share construction employment	10.52	6.63	0.00	48.68
Share hospitality employment	5.88	5.45	0.00	53.97
Share services employment	2.80	1.44	0.00	12.50
Share buildings from 1900-1920	1.91	3.52	0.00	97.77
Share buildings from 1921-1940	2.69	4.32	0.00	90.17
Share buildings from 1941-1950	3.31	5.48	0.00	99.92
Share buildings from 1951-1960	7.04	8.69	0.00	99.62
Share buildings from 1961-1970	12.73	12.99	0.00	100.00
Share buildings from 1971-1980	19.33	15.29	0.00	100.00
Share buildings from 1981-1990	15.81	14.31	0.00	100.00
Share buildings from 1991-2000	19.38	17.57	0.00	100.00
Car use index	26.66	19.32	-18.71	81.49
Pedestrian index	45.58	15.10	0.00	85.79
Building height index	61.20	79.36	-61.46	254.78
Neighborhood quality index	85.44	50.58	-25.18	226.65
Log distance to metro area center	1.58	1.01	-6.02	9.43

Variable	Average	Std. Dev.	Min.	Max.
Log distance to municipality center	1.65	2.72	-5.06	9.43
2008 Variables				
No population in 2008	0.02	0.13	0.00	1.00
Gravities				
Exits (from highways or roads)	7.76	152.00	0.03	22459.80
Winery	0.35	15.90	0.00	1961.40
ATM	2204.90	216575.00	0.11	36176200.00
Train Station	2.24	96.30	0.00	14831.50
Commuter Rail Station	37.40	1272.60	0.00	155200.20
Bus Station	1.65	59.20	0.00	5951.60
Ferry Terminal	0.05	1.29	0.00	147.10
Marina	2.73	215.30	0.00	32409.00
Public Sports Airport	0.02	1.55	0.00	247.00
Airport	0.05	0.65	0.00	74.00
Business Facility	4.61	705.30	0.00	119087.90
Grocery Store	623.60	12388.90	0.14	1150218.00
Automobile Dealership	570.10	45786.10	0.07	6990278.00
Petrol/Gasoline Station	929.40	131834.60	0.09	22218260.00
Motorcycle Dealership	1.03	55.30	0.00	6241.80
Restaurant	2176.80	68143.60	0.16	7465020.00
Nightlife	79.10	2241.80	0.01	252470.60
Historical Monument	33.90	1275.30	0.01	165075.90
Bank	2235.50	216704.70	0.11	36176200.00
Shopping	9.25	433.40	0.00	59529.20
Hotel	405.20	33195.30	0.08	5502869.00
Ski Resort	0.00	0.02	0.00	3.79
Other Accommodation	0.00	0.00	0.00	0.03
Tourist Information	2.20	99.60	0.00	13718.00
Rental Car Agency	34.10	4619.80	0.00	779683.20
Parking Lot	23.20	814.80	0.02	115492.80
Parking Garage/House	102.30	2870.50	0.00	400110.70

Variable	Average	Std. Dev.	Min.	Max.
Park and Ride	0.04	1.45	0.00	158.90
Auto Service and Maintenance	224.90	14561.00	0.05	2346258.00
Cinema	11.40	383.10	0.00	42780.50
Rest Area	0.11	0.84	0.00	92.40
Performing Arts	14.50	468.80	0.00	54141.10
Bowling Centre	0.33	14.90	0.00	1407.20
Sports Complex	5.53	392.30	0.00	59071.60
Park/Recreation Area	35.10	1280.10	0.01	170288.90
Casino	4.52	592.80	0.00	98100.60
Convention/Exhibition Centre	0.63	25.30	0.00	3156.50
Golf Course	1.80	149.10	0.00	23836.20
Civic/Community Centre	7.45	288.10	0.00	32318.00
Amusement Park	0.06	2.33	0.00	372.70
Sports Center	41.50	1481.70	0.01	151129.50
Ice Skating Rink	0.12	11.90	0.00	1974.10
Tourist Attraction	1328.40	216692.80	0.02	36589410.00
Hospital	2.54	76.50	0.00	11558.20
Higher Education	2.11	85.20	0.00	11387.40
School	209.40	11426.30	0.02	1833595.00
Library	1.23	79.40	0.00	13151.50
Museum	20.90	1247.30	0.00	184048.80
City Hall	3.31	133.10	0.00	19628.20
Police Station	12.50	394.60	0.01	52615.80
Post Office	81.00	9585.60	0.00	1611745.00
Department Store	0.85	28.10	0.00	4109.30
Home Specialty Store	0.02	0.19	0.00	11.20
Pharmacy	720.30	18937.70	0.06	2035484.00
Specialty Store	0.17	12.20	0.00	1693.20
Sporting Goods Store	0.30	24.00	0.00	3916.90
Medical Service	5.57	255.20	0.00	30171.50
Consumer Electronics Store	2.46	109.00	0.00	16685.20

Variable	Average	Std. Dev.	Min.	Max.
Industrial Zone	12.50	469.30	0.02	68646.40
Place of Worship	52.50	822.00	0.00	102681.30
Embassy	60.10	2752.60	0.01	365454.10
Book Store	67.00	4886.10	0.00	648498.10

Source: own elaboration on data from INE (2009), INE (2005) and ESRI (2009). Neighborhoods are squares with sides measuring 0.005 degrees. Distance to metro and municipality center is population weighted. Construction of indices and gravities detailed in the Appendix.

Table 4: Change in the Native Population vs. the Change in the Migrant Population (OLS)

Dependent variable	$\Delta nat_{k,m,2001\_2008}$					
Sample	All		Winsor			
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Non-Western Immigrants $\Delta im^{NW}_{k,m,2001\_2008}$	-0.517	-0.119	-0.272		-0.249	-0.296
	[0.020]***	[0.041]***	[0.040]***		[0.049]***	[0.076]***
Western Immigrants $\Delta im^{W}_{k,m,2001\_2008}$	2.127	1.332	3.125	2.691	3.797	4.012
	[0.325]***	[0.281]***	[0.582]***	[0.589]***	[0.804]***	[1.002]***
Eastern Europe				-0.266 [0.069]***		
Latin America				-0.425 [0.075]***		
Subsaharian Africa				-0.874 [0.193]***		
Maghreb				0.105 [0.125]		
Rest of Asia				0.010 [0.146]		
$f(pop_{k,m,2001})$ and Controls	No	Yes	Yes	Yes	Yes	Yes
Initial concentration	No	No	No	No	Yes	Yes
Population Average Weights	No	No	No	No	No	Yes
Adjusted R^2	0.179	0.329	0.422	0.428	0.426	0.535
Observations	28,521	28,521	22,041	22,041	22,041	22,041

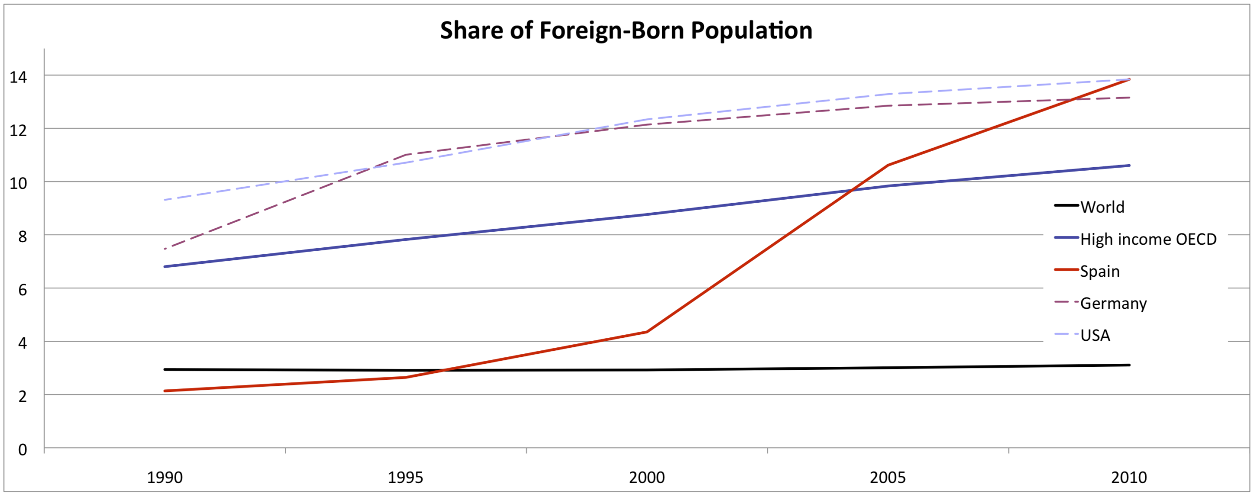
Source: own elaboration on data described in Table 3. Robust standard errors in brackets. ***: significant at 1%; **: significant at 5%; *: significant at 10%. $f(pop_{k,m,2001})$ refers to a quartic polynomial in $pop_{k,m,2001}$. Winsor means that the sample has been reduced by dropping the neighborhoods with the 1 percent lowest and highest population growth in 2001-2008 (weighted by their 2001 population). All specifications include metro area dummies. Initial concentration refers to the share of immigrants of each origin in 2001.

Table 5: Instrumental Variables Specifications

Dependent variable	$\Delta nat_{k,m,2001\_2008}$				$\Delta Buildings_{k,m,2001\_2008}$
Sample	All		$u \geq 11\%$	$u < 11\%$	All
Variables	(1)	(2)	(3)	(4)	(5)
Non-Western Immigrants $\Delta im^{NW}_{k,m,2001\_2008}$	-0.279	-0.323	-0.517	-0.408	0.427
	[0.055]***	[0.059]***	[0.082]***	[0.130]***	[0.195]**
Western Immigrants $\Delta im^{W}_{k,m,2001\_2008}$	0.652	1.067	3.492	-0.068	1.008
	[0.354]*	[0.752]	[1.609]**	[0.495]	[0.404]**
First Stage F-stat ($\Delta im^{NW}_{k,m,2001\_2008}$)	191.19	83.38	46.55	70.12	113.96
First Stage F-stat ($\Delta im^{W}_{k,m,2001\_2008}$)	7.81	12.66	15.62	3.35	41.78
Joint Wald F-stat	19.87	61.13	189.07	56.23	180.33
Observations	28,521	28,521	13,929	14,592	26,578

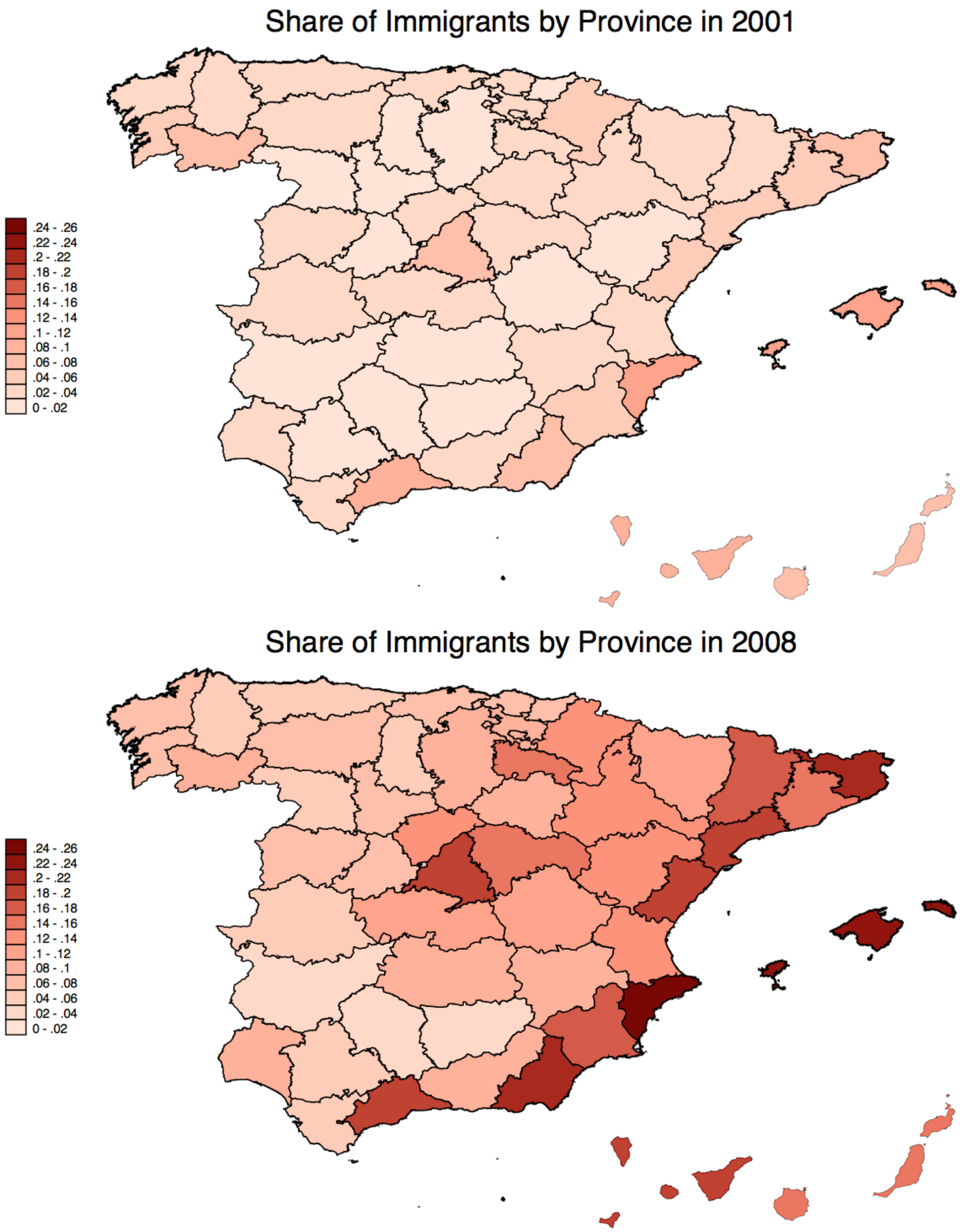
Source: own elaboration on data described in Table 3. Robust standard errors in brackets. ***: significant at 1%; **: significant at 5%; *: significant at 10%. Most developed and poor origins, as in Table 2. Controls in this table correspond to those in column (5) of table 4. Population weights are averages of their 2001 and 2008 numbers for columns (2) to (5). Regressions run using *ivreg2* command in Stata, written by Baum, Schaffer, and Stillman (2014).

Figure 1: Evolution of the foreign-born population share



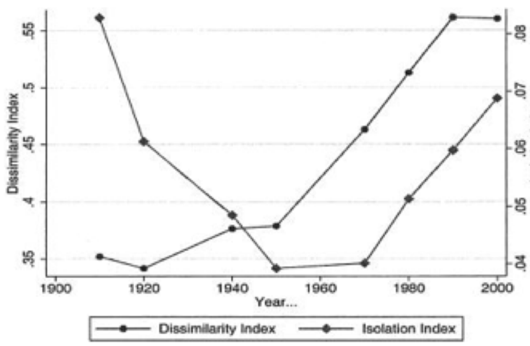
Source: own elaboration on data from World Bank (2013)

Figure 2: Distribution of the share of immigrants by province in 2001 and 2008

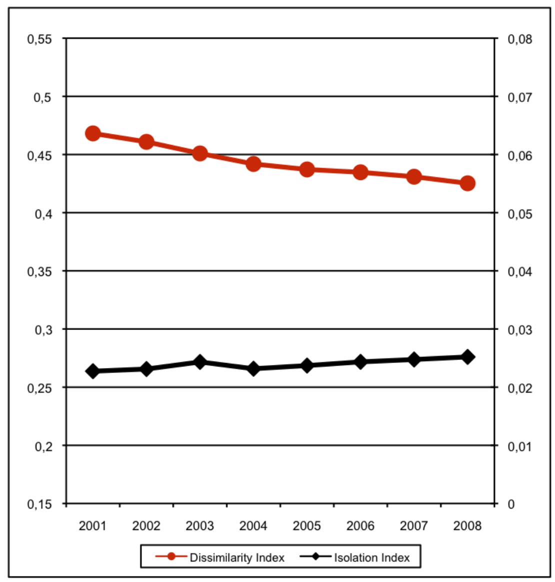


Source: own elaboration on data from INE (2009)

Figure 3: Comparing segregation levels in Spain and the US



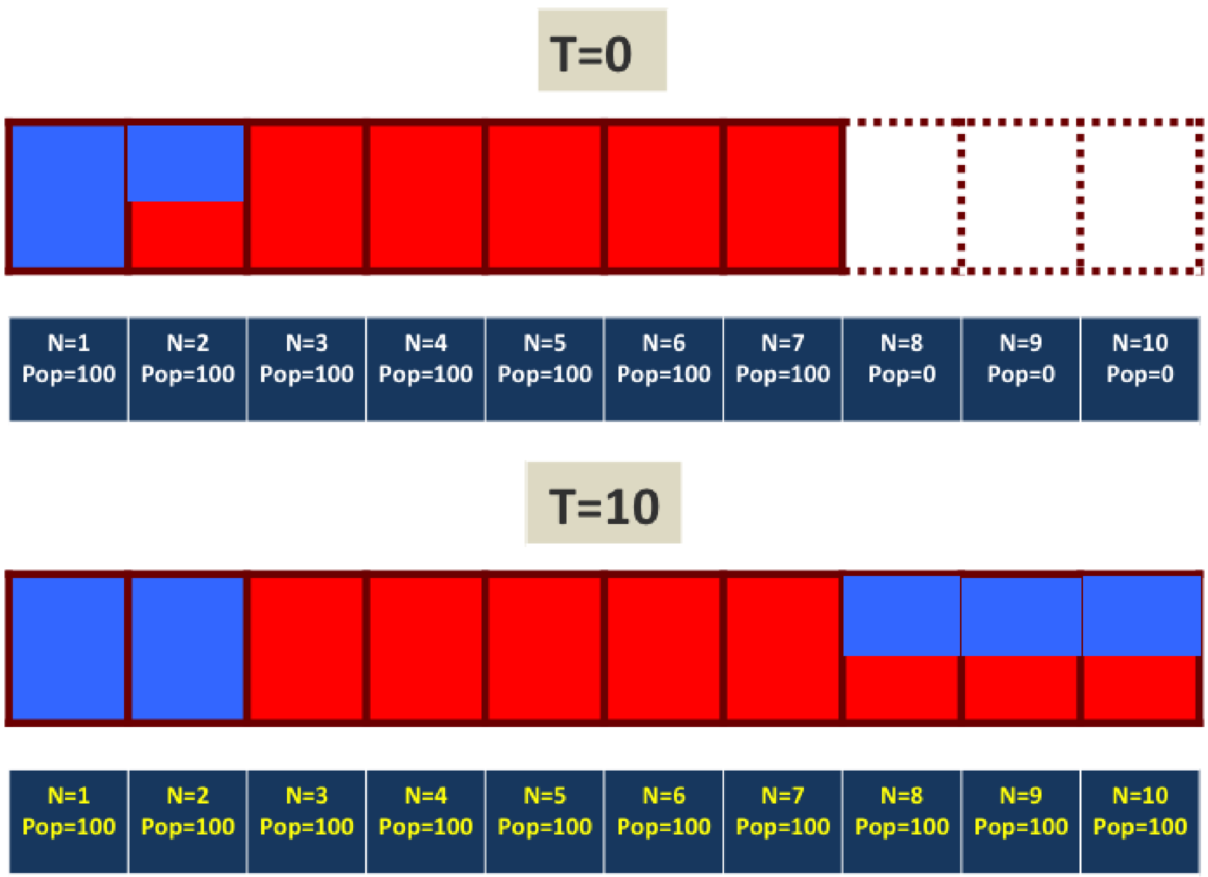
US



SPAIN

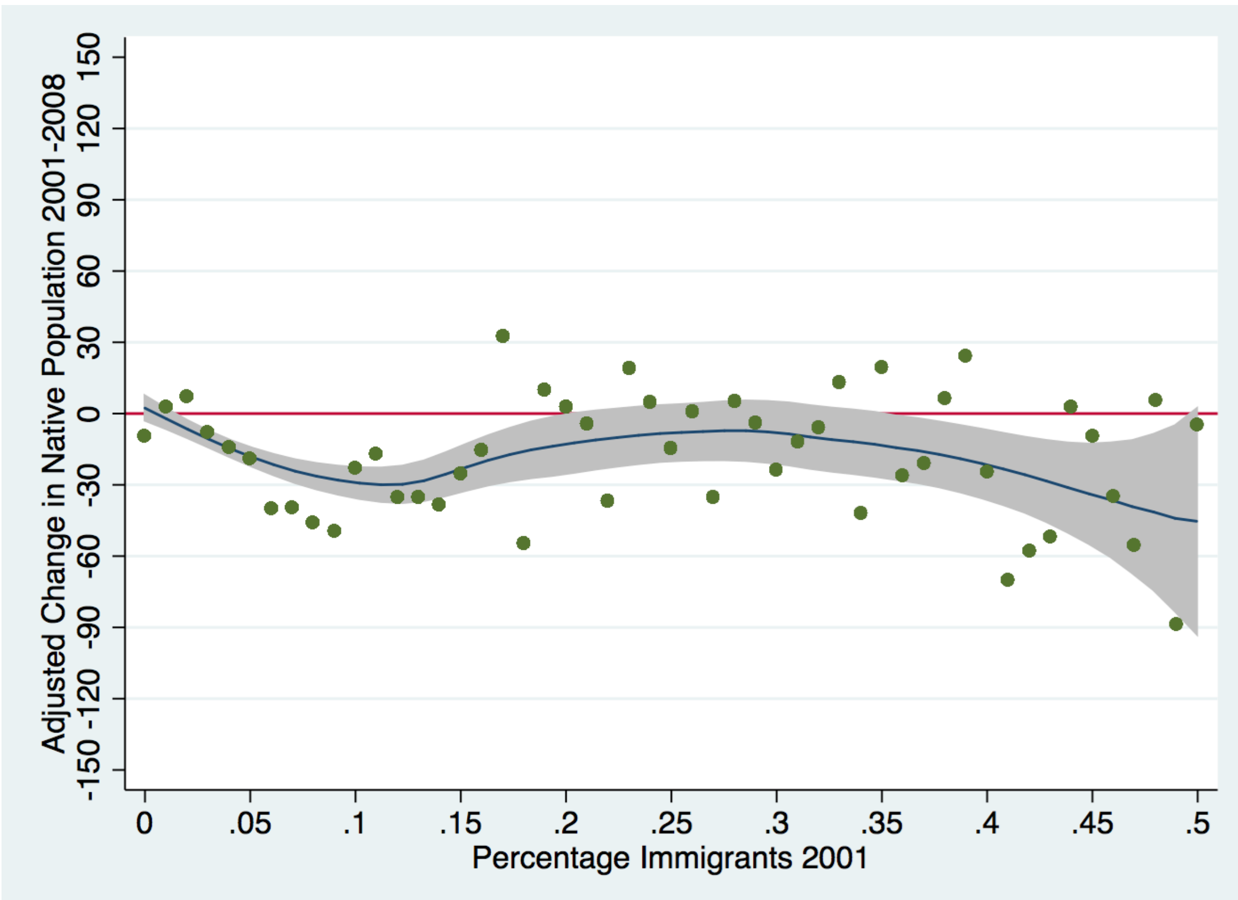
Source: Cutler, Glaeser, and Vigdor (2008) and own elaboration on data from INE (2009). In the Spanish case, neighborhoods are 2008 censal sections taken back in time using our geocoded addresses. Indexes calculated averaging immigrant communities with at least 1,000 inhabitants in a metro area, weighted by the size of the community in each metro area.

Figure 4: LATE of Neighborhood Ethnic Change Parameters



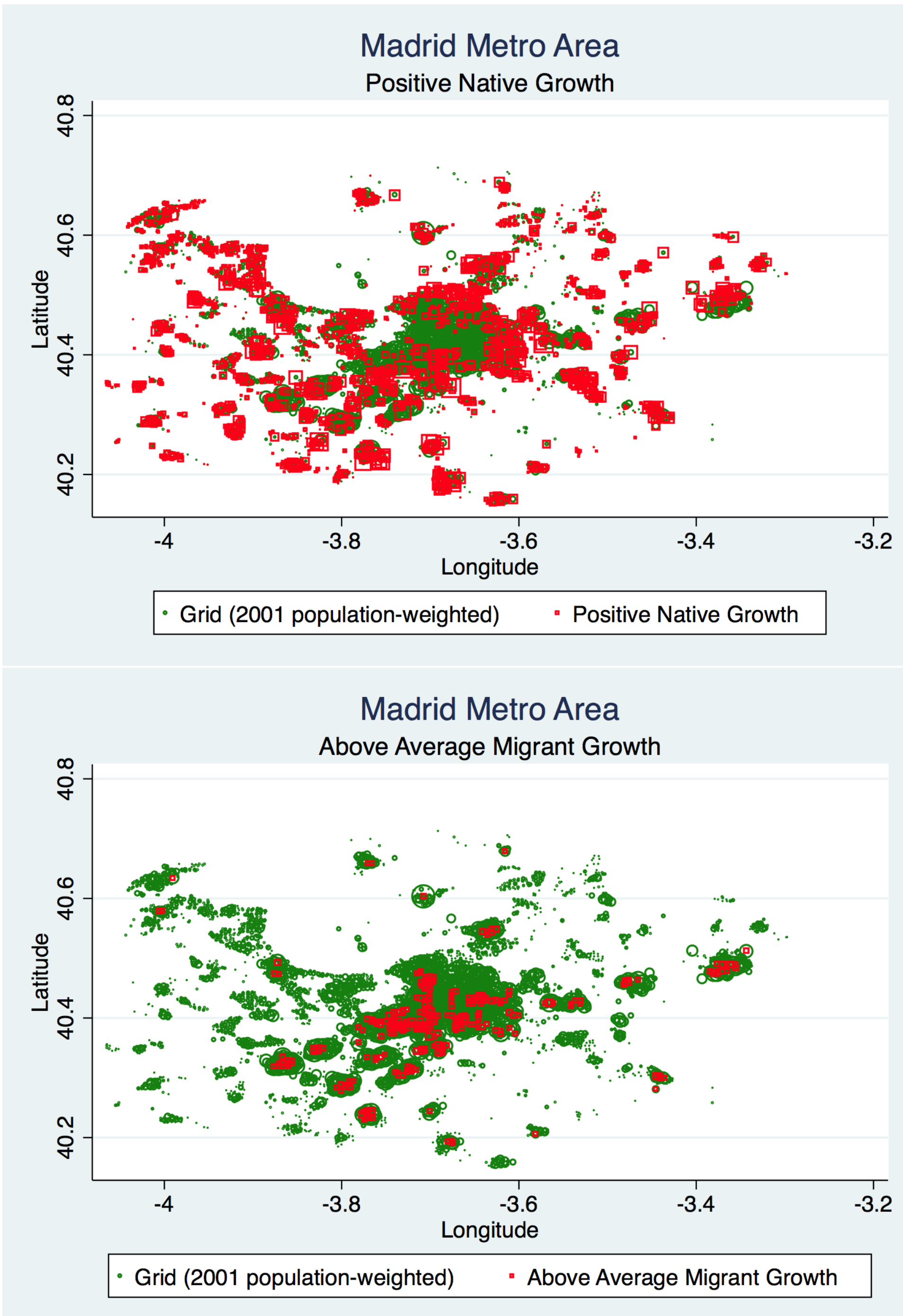
The figure illustrates a hypothetical city with 10 neighborhoods depicted as contiguous rectangles. Eight of the neighborhoods are populated at T=0, whereas 3 new neighborhoods are only developed subsequently at T=1. The share of immigrant ethnic minorities (majorities) in each neighborhood is signified by the area in in blue (red). By assumption, tipping dynamics are present in the older established neighborhood (N=2), which have crossed the tipping threshold at T=0. However, newly developed neighborhoods are seeing considerable ethnic mixing. In this example, empirical methods that use past settlement patterns to estimate future ethnic majority flows provide a local average treatment effect (LATE) that is starker than the global average treatment effect.

Figure 5: Local average smooth plot of adjusted native inflows (2001-2008) by initial share of immigrants (2001)



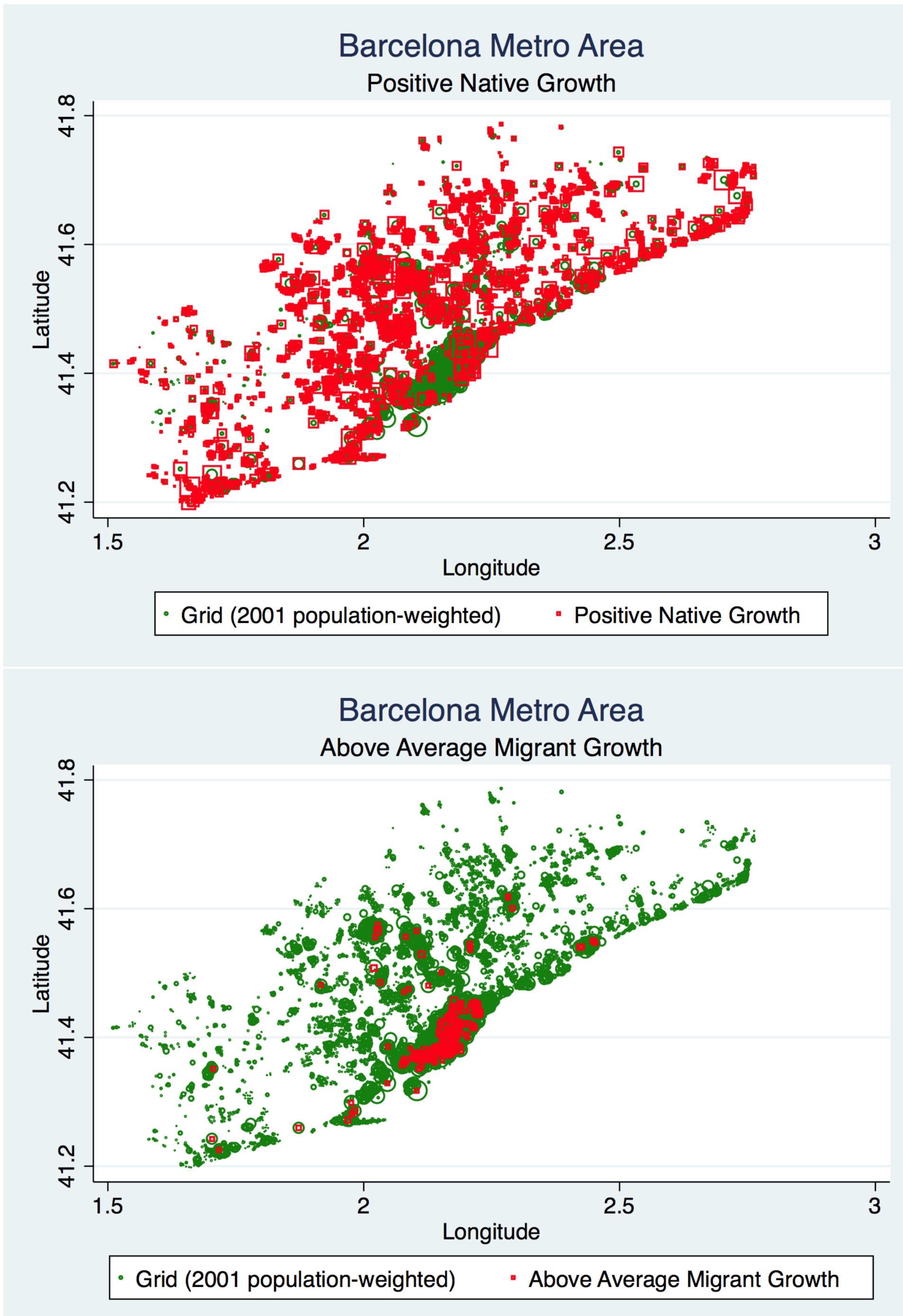
The graph uses information on all neighborhoods that were populated in 2001. It uses the Epanechnikov kernel to smooth local means of adjusted native inflows by group of neighborhoods sorted according to their initial immigrant concentration. Native inflows are adjusted by estimating a regression on all variables in table 4 and adding the residuals to the prediction corresponding to the national average neighborhood characteristics.

Figure 6: Neighborhoods with positive native growth and with above average migrant growth in Madrid metro area (2001-2008)



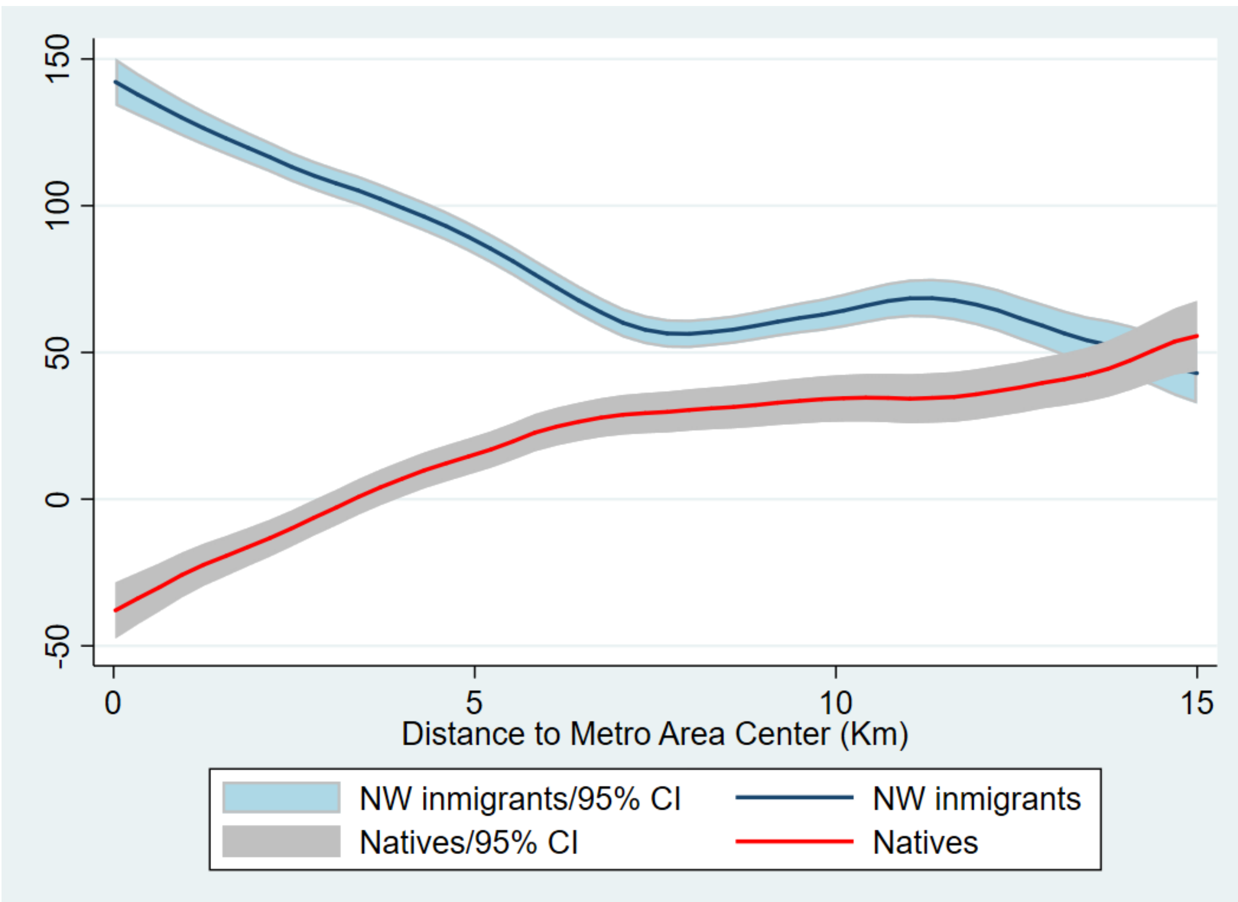
Source: own elaboration on data from INE (2009). Circles proportional to population in 2001. Squares proportional to 2001-2008 changes.

Figure 7: Neighborhoods with positive native growth and with above average migrant growth in Barcelona metro area (2001-2008)



Source: own elaboration on data from INE (2009). Circles proportional to population in 2001. Squares proportional to 2001-2008 changes.

Figure 8: Local average smooth plots of native and non-Western immigrant inflows by distance to city center



B Construction of the dataset

B.1 Sample selection

We start from 28,870 grids matched between 2001 and 2008 in the metro areas defined by Ministerio de Vivienda (2007) at the municipality level. We next drop 13 grids where only one individual was registered and he/she was aged less than 15 in 2001 (6 cases) or 2008 (7 cases). This leads us to 28,857 observations. 328 of them have been identified as having address problems either in 2001 or in 2008. This amounts to 0.6 percent of the relevant population in 2001. The most problematic places are Tenerife (10 percent) and Tarragona-Reus (5 percent). We are left with 28,529 valid observations. We still drop for our analysis 8 neighborhoods that are outliers either in terms of native population growth (2 neighborhoods that lost more than 8,000 natives in 2001-2008) or in terms of the predicted growth from the instrument (4 neighborhoods predicted to get more than 4,000 Western immigrants into them or around them and 2 neighborhoods predicted to get more than 8,000 non-Western immigrants). These outliers do not affect the results but some graphical representations in the paper worsen considerably when they are included. They contained overall 90,923 inhabitants in 2001 and 80,169 in 2008. After these operations, we are left with our main sample of 28,521 observations.

B.2 2001 Census Indices

The 2001 Spanish Census (INE, 2005) provides us with a set of variables referred to each of 34,251 censal sections in Spain. Censal sections are administrative divisions for electoral purposes and are supposed to have between 500 and 2,500 inhabitants. In 2001, their average population was 1,193 (s.d.=590). 94 percent of them had the correct size.

We assign each of our addresses to the 291 average characteristics of its censal section in 2001. The included variables and their descriptives are shown in Table 3. Most variables are self-explanatory except for the four summary indices that we calculate. These indices are the results of two principal component analysis on several variables.

- The first one uses the following Census variables: share of employed individuals in the neighborhood commuting by car alone, by car with passengers, by bus, by subway, by train, by motorcycle, by bike or on foot, share of population owning one car, two cars or three or more. We take the first two principal components weighted by 2001 population. The first accumulates 55 percent of the variance and we call it the car use index. The second gathers 16 percent of the variance and we call it the pedestrian index.

- The second one uses the following Census variables: share of population with public water, private water, sanitation services, gateway services, aspect of the building (scored by the interviewer), gas services, phone, hot water, heating, air conditioner, height of the building, elevator, noise level, contamination, crime, green areas and dirtiness. Different categories of these variables are included to total 42 variables. Again, we take the two principal components weighted by 2001 population. The first accumulates 38 percent of the variance and we call it the building height index. The second gathers 13 percent of the variance and we call it the neighborhood quality index.

B.3 2008 Gravities

We geocoded our Padrón (INE, 2009) data by matching each address with addresses from ESRI StreetMap Premium Europe NAVTEQ 2009 Release 2 (ESRI, 2009). We end up with 7,568,601 uniquely identified addresses. For each address, we calculated its distance in meters to a series of 62 features (points of interest) from the map server, such as hospitals, exit roads, schools, bus stops, metro stops, etc. In the end, for each address, we have six different measures of amenities for each of the 62 points of interest. The first measure is the minimum distance between each address and each of the points of interest. The other five measures are gravities: sums of points of interest in Spain weighted by distance. That is:

$$g_i^{p,\alpha} = \sum_{n_p=1}^{N_p} d_{i,n_p}^{-\alpha} \tag{13}$$

where i is an address, p is a point of interest (i.e. hospitals), N_p is the number of points of interest p in the radius where i is located, α is a coefficient that takes values $\{0.5; 1; 2; 3; 4\}$ and d_{i,n_p} is the distance between address i and point of interest n_p (i.e. one particular hospital).

Table 3 shows summary statistics for gravities calculated for $\alpha = 2$. They have been multiplied by 1,000,000 to make the numbers visible.