Immigration and the Neighborhood: Online Appendix

Albert Saiz, Susan Wachter

Abstract

In this Online Appendix (OA) we present the theoretical model that guides the empirics, extensions to the results in the paper, and a more extensive discussion of the empirical identification strategy. We also display maps of the process of immigrant diffusion in the DC area.

1 Immigration, Native Mobility and Housing

We propose a basic framework based on conventional racial segregation models (Bailey, 1959; Schelling, 1971; Yinger, 1975; Courant and Yinger, 1975; Kanemoto, 1980). The model is the simplest that manages to illustrate well the main issues at play, both locally and in city-wide general equilibrium. We make a new modeling assumption of interest by considering income heterogeneity in the native population. We assume a city with an exogenously given native population of measure one. Among natives, income has a uniform distribution so that a measure N of inhabitants has income equal or below $\chi + N$, where χ is the minimum income (maybe a government transfer) and $N \in [0, 1]$. Immigrants tend to cluster in specific city neighborhoods. In the 1980s, 95 percent of the change in the number of immigrants (75 percent in the 1990s) was concentrated in a number of census tracts that corresponded to about 25% of the 1980 metropolitan US population. We therefore assume that there are four otherwise identical neighborhoods and that immigrants tend to concentrate in one of them (neighborhood 4) because of ethnic-composition preferences. Nevertheless, if natives have preferences for living with other natives, the only stable equilibria in the housing market will *endogenously* imply clustering of immigrants without prior assumptions. The utility function of native i is Cobb-Douglas in consumption (C_i) and the share of natives in the neighborhood where i resides (ϕ_i): $U_i = C_i^{\rho} \cdot \phi_i^{(1-\rho)}$. Each person consumes an identical unit of housing. Housing supply is assumed to be linear in rents: $R_k = \beta \cdot Pop_k$, where Pop_k is the total population in neighborhood k. Consumption depends on income and rents so that $C_i = \chi + N - R_i$, where R_i is the rent in the location chosen by the individual. In this simple model all houses are of the same quality and house values are directly proportional to rents, capitalizing their present discounted value at the discount rate d: $price_k = \frac{R_k}{d}$

Without immigration, all equilibria in the residential market imply that the population is evenly spread throughout each of the neighborhoods. If population (and thus rents) were lower in one of the neighborhoods, everyone would like to move there. There are multiple all-native equilibria with different income mixes by neighborhood.

With immigration and native preferences for segregation, the equilibrium in the housing market implies that the poorest natives will live in the immigrant neighborhood, since richer individuals have a higher willingness to pay for segregation. The rest of the native population will be evenly distributed in the three other neighborhoods. In a "mixing" equilibrium there is a marginal native with income $\chi+\underline{N}$ who is indifferent between the immigrant neighborhood and the rest of the city:

$$(\chi + \underline{N} - \beta \cdot [F + \underline{N}])^{\rho} \cdot \left(\frac{\underline{N}}{\underline{N} + F}\right)^{(1-\rho)} = \left(\chi + \underline{N} - \beta \cdot \left[\frac{1 - \underline{N}}{3}\right]\right)^{\rho} \tag{1}$$

Where F is the number of foreign-born individuals. The native share in the immigrant-dense neighborhood can be expressed as $\phi = \left(\frac{\underline{N}}{\underline{N}+F}\right)$.[1] Equation (1) implicitly defines the number of natives in neighborhood 4 ($\underline{N}$) as a function of the number of immigrants. Differentiating this equation with respect to F yields:

$$\frac{\partial \underline{N}}{\partial F} = \frac{-\beta \cdot \phi^{\frac{1-\rho}{\rho}} \cdot \underline{N} - \phi \cdot C_{NAT} \cdot \left(\frac{1-\rho}{\rho}\right)}{\left(1 + \frac{\beta}{3}\right) \cdot \underline{N} - (1-\beta) \cdot \phi^{\frac{1-\rho}{\rho}} \cdot \underline{N} - (1-\phi) \cdot C_{NAT} \left(\frac{1-\rho}{\rho}\right)} \tag{2}$$

Where $C_{NAT} = \left(\chi + \underline{N} - \beta \cdot \left[\frac{1-\underline{N}}{3}\right]\right)$. This expression is generally negative for equilibria with some ethnic mixing. To see an example of that, assume that the initial level of immigration is zero (and thus $\phi = 1$) to obtain:

$$\left.\frac{\partial \underline{N}}{\partial F}\right|_{\phi=1} = \frac{-\beta \cdot \underline{N} - C_{NAT} \cdot \left(\frac{1-\rho}{\rho}\right)}{\frac{4}{3} \cdot \beta \cdot \underline{N}} < 0 \tag{3}$$

More generally $\left.\frac{\partial \underline{N}}{\partial F}\right|_{\rho_{\leq 1}} < 0$, i.e. *there is native flight out of the immigrant neighborhood, if natives display preferences for segregation* ($\rho \leq 1$) and there are any natives remaining.[2]

[1] With extremely high tastes for segregation or major immigration inflows, there may not be an equilibrium with a marginal native (i.e., the model may tip toward total segregation). However, the income effect typically helps to achieve some mixing: as the immigrant population in the immigrant neighborhood increases, the number of natives decreases, but the marginal native is poorer, and thus has a lower ability to pay for segregation. Since some low-income individuals do not have the income to respond to their tastes for segregation by moving to all-native neighborhoods, they may actually display stronger preferences for immigration limits or voice stronger opposition to immigration through their political choices, or in opinion surveys and daily behavior.

[2] Although the model does not have a closed form solution, simulations (unreported in the paper for brevity, but available on request) were used to generalize these "native flight" results for combinations in the main parameters of interest.

How about relative rents/values? In the situation where natives are indifferent about the ethnic composition of their neighborhoods ($\rho = 1$), and without massive levels of immigration (this is with $F \leq \frac{1}{3}$) we have $\left.\frac{\partial R_j}{\partial F}\right|_{\rho=1} = \left.\frac{\partial R_4}{\partial F}\right|_{\rho=1} = \frac{\beta}{4}$. Hence, home values will increase in all neighborhoods equally if natives are indifferent about the ethnic composition of the neighborhood; even when *immigrants* exhibit a preference for clustering together. In this scenario, we should expect a zero correlation between immigration and neighborhood values.

With $\rho < 1$ (native preferences for ethnic homogeneity), in an interior equilibrium, housing values should grow more slowly in the immigrant areas: $\left.\frac{\partial R_j}{\partial F}\right|_{\rho<1} > \left.\frac{\partial R_4}{\partial F}\right|_{\rho<1}$; the growth in the immigrant share of the destination neighborhood is associated with native flight of relatively high-income individuals. However some low-income individuals have an incentive to remain in the immigrant neighborhoods due to the compensating differential offered by the lower housing rents.[3] This implies that the changes in home values associated with immigrant inflows may correspond to the neighborhood valuations of the relatively low-income individuals who are at the relevant margin. To illustrate these effects, in OA Figure 1, we present the results of simulations of the model with parameters $\beta = 1$, $\rho = 0.9$, and $\chi = 0.5$. Rents (and hence values) are growing everywhere with increasing immigration levels. Nevertheless, the rate of growth is faster in the *native* neighborhoods.

Note that *very large* immigrant inflows could fully displace any remaining marginal natives from the immigrant neighborhood.[4]. In a fully-segregated scenario, further immigration inflows involve *growing values* in the enclave and no price inflation in the rest of the city. Note also that if natives exhibit a preference for diversity ($\rho > 1$), values (and population) will go up in the immigrant neighborhood. An empirical positive association between changes in values and changes in immigrant density is therefore hard to interpret. However, a negative association (controlling for other location and housing quality attributes) provides an *unequivocal sign of native preferences for segregation*. Intuitively, a non-arbitrage

[3] To see that, remember the equilibrium condition $(\chi + \underline{N} - R_j)^{\varphi} = (\chi + \underline{N} - R_4)^{\varphi} \cdot \phi^{1-\varphi}$, $\forall j \neq 4$. With $(\phi < 1)$ then $(\chi + \underline{N} - R_j)^{\varphi} < (\chi + \underline{N} - R_4)^{\varphi}$, which implies $R_j > R_4$. Note that with a very high distaste for diversity among natives, price growth in immigrant areas might even be negative in absolute terms despite the fact that the average city rent growth is positive.

[4] This is an conceptual matter that does not concern the empirics, since we are using *changes* in inmigrant density empirically and controlling for lagged immigrant density

condition ensures that prices cannot be lower in a location unless there is a perceived negative compensating differential: otherwise opportunistic natives move in until the price gap is bridged.

Note that, in all cases, the model shows that immigration will push *average metropolitan* housing values up, consistent with the recent empirical literature (Saiz, 2003, 2007; Ottaviano and Peri, 2006). Even with tastes for segregation, values may increase in immigrant neighborhoods (this depends on the parameters of the model and on immigration levels), but not as fast as *relative to* the rest of the metropolitan area.

2 OLS Robustness Tests

2.1 Unobserved Structural Quality Attributes

The results in paper's Table 1 show a robust negative association between changes in housing values and growth in immigrant density. A potential interpretation of the results involves a housing *filtering* story, where immigrants do not make substantial investments in their housing units. This story does not require negative capitalization effects on land values. Given the magnitude of our estimates, this would imply a physical depreciation of immigrant-occupied homes substantially greater than 25 percent each decade.[5] The fact that median home values change also in neighborhoods where the median owner is a native makes this hypothesis less likely. Nevertheless, despite the fact that we do include controls for changes in quality in our regressions, some quality attributes may remain unobservable in Table 1.

To address this issue, we use data from the American Housing Survey. The 2001, 2003, and 2005 issues of the survey include information about the foreign-born status of household members in the sample. The data also contain detailed information on housing quality and investments in renovation, maintenance, alterations, and repairs at the household level. We run regressions where housing investment (OLS regression) and up to 17 quality indicators (logit specifications) are on the left-hand side and an indicator that takes a value of one if

[5] Land values are typically high in these areas, so the structure accounts for a relatively smaller fraction of the house value. If the impact of immigration on the total price has to come from changes in the value of the structure, this implies a much higher depreciation rate on the physical structure.

any of the household members is foreign born is the main explanatory variable (presented in Online Appendix Table 1).[6] For each quality/investment attribute, we run regressions that only control for year fixed effects (left columns) or for a more complete set of household attributes: income, marital status, gender, age of the reference person, and a dummy for recent movers (right columns). We present regressions that use the cross-sectional variation in the pooled data from 2001, 2003, and 2005 (upper rows), and regressions that include housing unit fixed effects for those observations appearing both in the 2001 and 2005 samples (bottom rows). Since we use two time observations, this fixed effects model is identified off changes in the immigrant status of the homeowner.[7]

The evidence, in OA Table 1, shows that immigrant homes are not of lower quality, and some of their attributes may even be better (significant coefficients at the 5% level are highlighted). Moreover, in no case is a *change* toward immigrant ownership of a housing unit associated with negative changes in observable quality.

Quality is a stock variable and may evolve very slowly. But the total expenditure on maintenance and renovation is a flow variable that is under direct control of the household. The evidence on this variable does not support either the view that immigrant homeowners depreciate faster their housing assets faster by investing less in maintenance and renovation.

In sum, changes in structural housing quality in the homes that are taken over by immigrants cannot account for the results in Tables 1 and 2.

2.2 Unobserved Neighborhood Characteristics or Trends

In Online Appendix Table 2, we run several robustness tests to Table 1 in the paper. Column 1 shows the results for the full sample of *all US metropolitan* areas in the 80s and 90s. In column 2, in order to focus on changes in values in *new* immigrant neighborhoods, we restrict our sample to those tracts with initial immigrant densities below the MSA median. In column 3, we add two indicators of the environmental quality of the neighborhood: the shares of

[6] We obtained similar results using the foreign-born status of the reference person instead.

[7] Changes in quality and immigrant status of the homeowner between two consecutive sample years are noisier, and therefore we do not include the 2003 observations, relying rather on the "long-differences" in the variables.

area in the tract covered by water and devoted to industrial or commercial uses in 1992. Table 2, column 4 uses the log of *median* house value as dependent variable. We only have median home values by census tract for 1990 and 2000, so we restrict our attention to the 1990s. Our baseline estimates are remarkably insensitive to all these specification changes.

In column 5 we address potential issues concerning heterogeneity in housing supply elasticities or potential idiosyncrasies in the geographic location of immigrant communities.[8] In Table 1 we controlled for past density (in unreported regressions, controlling for central city location did not change the results). We go further now and divide the sample of neighborhoods into quartiles defined by density and distance to the central business district the decade before *within each MSA*. We then run separate regressions as in table 1, column 2, within each of the 16 possible density-distance quartile combinations. Finally, we average the individual results using the number of tracts in each of the 16 resulting groups as weights. The *average* result is notably similar to the ones in previous specifications. In fact, negative relative associations between immigration and values were found within *each of all* these 16 very different types of communities (results available on request): from dense areas close to the city center, to low-density suburban locations far away from the metro core. These results are not consistent with a sorting story where immigrants are moving into specific types of locations, but they are consistent with a treatment effect of immigration within each of these different types of areas.

In the last column of Table 2, we include lagged immigrant density. Again, we want to control for general trends in amenities and housing values in the areas where immigrants tended to settle in the past.[9]

[8] If the law of one price holds at the MSA level, housing prices should reflect the valuation of the neighborhood's attributes by the marginal mover regardless of the elasticities of housing supply (to see this, simply assume heterogeneous housing supply elasticities in the model in section 2).

[9] In unreported specifications we also conducted separate regressions for each of the available decades (1980s, and 1990s). The relative association of the change in the foreign born and housing price inflation was negative in both decades.

3 Choice of the Spatial Decay Parameter (β)

In the text we defined $Pull_{i,T} = \sum_{\substack{j \neq i \\ j \in M}} \frac{(ISH_{j,T-10}) \cdot Area_j}{(d_{ij})^{\beta}}$. $Pull_{i,T}$ is our estimate of the immigrant "geographic gravity pull" of a neighborhood i (which is located in a metropolitan area M) at time T. $(ISH_{j,T-10})$ is the share of immigrants in neighborhood j in the previous census (ten years ago), $Area_j$ is the area (square miles) of the corresponding jth census tract, and d_{ij} is the Euclidean distance in a longitude-latitude degree two-dimensional plane between neighborhoods i and j. Our measure of "gravity" is a weighted average of lagged immigrant densities in neighboring communities, where the weights are directly proportional to the area of neighboring tracts and inversely proportional to their distance from the relevant neighborhood. In order to choose the parameter β, we simulate different patterns of lagged spatial correlation in the distribution of immigrants in the 2000 census. For each potential β, we fit the model:

$$(ISH_{i,2000,M}) = A_M + \gamma \cdot Pull_{i,2000,M} + \varepsilon_{i,2000,M} \tag{4}$$

M is a subscript for metropolitan areas and A is a metro area fixed effect. We search for the parameter β that maximizes R-squared in equation (4): this is we choose β in order to maximize the correlation between immigrant shares in 2000 and neighboring immigrant density in 1990. The results from this exercise can be appreciated in Online Appendix Figure 2. There is a clearly concave relationship between β and the fit of our lagged spatial correlation model. The maximum predictive power of the model is obtained for a spatial decay parameter close to 1.6, which is the number that we settle for. However, the results in the paper are not sensitive to reasonable.

4 Wald Estimators and Identification with Geography

Consider the example in the paper with two types of neighborhoods. Recall that neighborhoods of the type C are close or contiguous to existing immigrant enclaves, whereas type F consists of neighborhoods located far from the immigrant enclaves. Denoting with an upper tilde variables that are partialed-out of the rest of explanatory variables in equation (**??**)

and dropping MSA and time subscripts for simplicity, the main equation of interest is:

$$\widetilde{\Delta \ln P_i} = \lambda \cdot \Delta \widetilde{ISH}_i + \xi_i \tag{5}$$

Since we are not certain that $cov(\Delta \widetilde{ISH}_i, \xi_i) = 0$, we can use the empirical knowledge that immigrant enclaves tend to expand to contiguous neighborhoods to add the following immigration equation:

$$\Delta \widetilde{ISH}_i = \delta \cdot D_i^C + u_i \tag{6}$$

Where D_i^C denotes a dummy variable that takes value one if neighborhood i is contiguous to an immigrant enclave and zero if it is located far away. Given this model, under the assumption $cov(D_i^C, \xi_i) = 0$, an unbiased estimator of λ is the Wald estimator:

$$\widehat{\lambda_{Wald}} = \frac{\frac{\sum \widetilde{\Delta \ln P_i} \cdot D_i^C}{N_C} - \frac{\sum \widetilde{\Delta \ln P_i} \cdot D_i^F}{N_F}}{\frac{\sum \Delta \widetilde{ISH}_i \cdot D_i^C}{N_C} - \frac{\sum \Delta \widetilde{ISH}_i \cdot D_i^F}{N_F}} \tag{7}$$

N_c and N_F denote the number of neighborhoods that are close and far from the immigrant enclaves (respectively), and $D_i^F = (1 - D_i^C)$. Note that in $\widehat{\lambda_{Wald}}$, D_i^C is used as an instrument for $\Delta \widetilde{ISH}_i$.

A potential caveat of this *naive* instrumental variable approach hinges on the exogeneity assumption of $Pull_{i,T}$. It is certainly possible that previous immigrants were attracted to neighborhoods with characteristics that were *becoming* relatively less valuable to natives, and which are also spatially correlated.

In the two-neighborhood-type world, this can be modelled as a direct effect of proximity to the enclave on subsequent changes in housing prices:

$$\widetilde{\Delta \ln P_i} = \lambda \cdot \Delta \widetilde{ISH}_i + \pi \cdot D_i^C + \xi_i \tag{8}$$

If $\pi \neq 0$, the previous exclusion restriction is no longer valid. However, the impact of proximity to an immigrant enclave is heterogeneous, which can be used to generate plausible new exclusions restrictions. For instance, consider now the existence (ex post) of two types of cities: cities with high immigration shocks, and cities with low-immigration shocks . It is a plausible (testable) proposition that immigrant enclaves in high-immigration cities are likely to expand more. We can now use a two neighborhood, two-city model and express the immigration equation thus:

$$\Delta \widetilde{ISH}_i = \delta_1 \cdot D_i^C + \delta_1 \cdot D_i^C \times D_i^H + u_i \tag{9}$$

Here D_i^H stands for a dummy variable that takes value one if the neighborhood is in a city with a high immigration shock. Note that variables are already partialed out from MSA fixed-effects (i.e. D_i^H is also controlled for). In the new system of equations. Now $(D_i^C \times D_i^H)$ can be used as an instrument for $\Delta \widetilde{ISH}_i$ in equation (8), in which we can control explicitly for both MSA fixed effects (price trends in high versus low immigration cities) and proximity to the enclave, as captured by D_i^C. Under the assumption $cov\left(D_i^C \times D_i^H, \xi_i / D_i^C, D_i^H\right) = 0$,[10] an unbiased estimator of λ is the following "differences-in-differences" Wald estimator:

$$\widehat{\widehat{\lambda_{Wald}}} = \frac{\left(\frac{\sum \Delta \widetilde{\ln} P_i \cdot D_i^C \cdot D_i^H}{N_{CH}} - \frac{\sum \Delta \widetilde{\ln} P_i \cdot D_i^F \cdot D_i^H}{N_{FH}}\right) - \left(\frac{\sum \Delta \widetilde{\ln} P_i \cdot D_i^C \cdot D_i^L}{N_{CL}} - \frac{\sum \Delta \widetilde{\ln} P_i \cdot D_i^F \cdot D_i^L}{N_{FL}}\right)}{\left(\frac{\sum \Delta \widetilde{ISH}_i \cdot D_i^C \cdot D_i^H}{N_{CH}} - \frac{\sum \Delta \widetilde{ISH}_i \cdot D_i^F \cdot D_i^H}{N_{FH}}\right) - \left(\frac{\sum \Delta \widetilde{ISH}_i \cdot D_i^C \cdot D_i^L}{N_{CL}} - \frac{\sum \Delta \widetilde{ISH}_i \cdot D_i^F \cdot D_i^L}{N_{FL}}\right)} \tag{10}$$

Where $D_i^L = (1 - D_i^H)$ and N_{CH} stands for the number of neighborhoods that are close to the enclave (C) and in high-immigration-shock (H) cities (same logic applies for other N subscripts in neighborhoods that are far from the enclave, F, and in low-immigration-shock areas, L). The numerator in this expression is a diffs-in-diffs expression of the change in log housing values across neighborhood types (close versus far to enclave) and city types (high versus low immigration shock). The denominator adjusts for the treatment probability

[10] In our view, we need rather convoluted ad hoc stories to expect $cov\left(D_i^C \times D_i^H, \xi_i / D_i^C, D_i^H\right) < 0$ a priori. Such a story involves immigrants choosing to move to cities where the areas close to enclaves (but not the enclaves themselves, since we are controlling for the share of the foreign-born and $Pull_{i,T}$) are declining. The literature on metropolitan migration has established that *relative* immigrant inflows at the metropolitan level are not sensitive to economic conditions and can be very well predicted by previous patterns of settlement at the MSA level, and national immigration levels (e.g. Card and , Saiz, 2007). The main preoccupation of the literature is that immigrants may be attracted to economically thriving (not declining) areas.

(change in immigration density) across groups. $\widehat{\lambda_{Wald}}$ combines the intuitions of diffs-in-diffs and Wald-IV estimators.

5 Graphic Intuition for Identification Approach

The spatial intuition for this geographic diffusion approach can be seen in Online Appendix Figure 3. The grids in the figure represent census tracts in a metropolitan area. Immigrant density is signified by darker coloring. At time T-10, census tract A is surrounded by immigrant-dense neighborhoods. Tract B is further from the areas of immigrant settlement, and C is further yet. At time T (after 10 years), assuming that the city is receiving further immigrant inflows , we would expect tract A to receive a higher immigrant intake.

Empirically, we will solely use the *heterogeneity* in the predictive power of the geographic diffusion as our effective source of identifying variation. $Pull_{i,T}$ may be a worse predictor of future immigration in neighborhoods that are already heavily immigrant. For example, if a large percent of the population in a tract is already composed of immigrants, proximity to other foreign-born areas will not likely be predictive of increases in its immigrant density. Conversely, mostly native neighborhoods that are close to the immigrant enclaves are "ripe" for increases in immigrant density. We model the fact that *geographic diffusion of immigration is more likely to go from more immigrant-dense neighborhoods to less immigrant-dense neighborhoods* by interacting $Pull_{i,T}$ with the lagged share of the foreign born. The intuition behind this strategy can be seen in Online Appendix Figure 4. Tracts A and B are exposed to a similar "geographic immigrant pull" in period T-10. However, we might expect immigration density to grow faster in tract B, since tract A is already more immigrant dense, and B is further from its steady-state equilibrium.

We use the general MSA level of immigration similarly. If there is no new immigration into the city, we would not expect the "gravity pull" of a neighborhood to be a particularly good predictor of future changes in the immigrant share. Therefore, the interaction between $Pull_{i,T}$ and the relative magnitude of immigration by metropolitan area is likely to improve the predictive power of the geographic diffusion model.[11] This research design can be grasped

[11] We divide the number of new immigrants in an MSA by its initial population to obtain the relative size

from Online Appendix Figure 5. At time T-10, tract A1 (in city 1) and tract A2 (in city 2) are identical in terms of proximity to immigrant neighborhoods. But since new immigration is greater in city 1, we can expect our geographic diffusion model to predict more immigration in A1 than in A2.

6 An Example: Metropolitan DC in the 90s

In Online Appendix Map 1, we display the level of immigrant density (immigrants divided by population) in the metropolitan tracts surrounding the District of Columbia. This was a large immigrant area in the 90s, with relatively new immigration inflows, and should serve to illustrate the ideas implicit in the "spatial diffusion" approach. The different shades of gray indicate each one of the tract groups divided in thirds of 10 percent (3.33%) by immigrant density (minimum density is 0%, maximum density is 60.20%). The darker the coloring the higher the initial immigrant density in 1990. Census places are also displayed for geographic identification. As can be seen, there were 3 major immigrant clusters: north of Garret Park, MD (North West of the City), East and around of Baileys Crossroads, VA (southwest area of DC), and North and East of Tacoma Park, MD (North East of the city). In map 2, we display a gradient of the estimated immigrant "pull" using the 1990 data. As can be seen, the areas close to the city in a clockwise semicircle from 7am to 2pm, with an added arm that extended northwest to Gaithersburg, were predicted to be more attractive to immigrants in the 90s. Indeed, map 3 in the OA shows that most of the areas that experienced significant growth in the immigrant share were in or around tracts with high estimated values of $Pull_{i,T}$. Note, however, that the previously described centers of immigrant settlement, with very high immigrant densities in 1990 already, actually experienced small changes in immigrant densities in the 90s. As we model via the interaction in OA Table 2, a high $Pull_{i,T}$ tends to be associated with increasing immigration densities, but less so in areas that were already heavily immigrant-dense.

of immigration.

7 IV: First Stage and Robustness

In Online Appendix Table 3 we present results from the first stages in the 2SLS estimation presented in Table 1. As can be seen, immigrant geographic pull is always a positive predictor of subsequent immigrant density growth. In column 2, we show that the contagion model works, as predicted, better in neighborhoods that were close to immigrant enclaves but not part of the enclave themselves (contagion goes from more dense to less dense areas), and in cities with more immigrants.

In table OA 4 we conduct several robustness tests pertaining to the baseline IV specification (as presented in Table 4, column 3). In Table 5, column 1, we include a third order polynomial in $Pull_{i,T}$ in the list of controls (the highest order polynomial that enters significantly in the specification). The concern here is that the effect of potential omitted local amenities correlated with the distance to existing enclaves may be very nonlinear. Results, however, do not change.

Another potential concern is that the $Pull_{i,T}$ variable may be correlated with *changes* in the immigrant density of *neighboring communities* and that there are spillovers across neighborhoods. In Table 5, column 2, we show that controlling for $Pull_{i,T+10} - Pull_{i,T}$ (a measure of the change in immigrant concentration in neighboring communities) on the right-hand side does not change the results either.[12]

In column 3, we present further evidence consistent with the "contagion" effects are indeed due to immigrants, as opposed to other factors. For instance, it could be that part of a city becomes run down, and that people in *and around* that part of the city start to leave, which reduces prices further and may attract *future* immigrants. We explore this hypothesis by using an approach based on the spatially-temporally-lagged evolution of prices. Concretely we calculate $LAGP_{i,T} = \sum_{\substack{j \neq i \\ j \in M}} \frac{\Delta \ln P_{T-10} \cdot Area_j}{d_{ij}^{\beta}}$. $LAGP_{i,T}$ is a gravity-based measure of the lagged evolution of housing prices in neighboring communities. For instance, a neighborhood may have experienced an amenity shock during the 80s, and that reduced prices and attracted immigrants. Then in the 90s, the amenity shock could have spilled over to neighboring low-immigration tracts. In Table 5, column 3 we control explicitly for the

[12] I am grateful to Vernon Henderson for suggesting this robustness test..

lagged evolution of prices in neighboring tracts, while still using the instruments based on proximity to lagged immigrant density. The results are almost unchanged, which refutes the alternative hypothesis of "spreading decline." In fact, there is some evidence of mean reversion with respect to neighboring community prices also: tracts that were close to areas that had been declining in the 80s tended to experience better-than-average performances in the 90s (controlling for our baseline set of 44 controls)..

In Table 5, column 4, we focus on the triple interaction between immigrant "pull," lagged immigrant shares, and metro immigration levels. Concretely, we add the triple interaction to the list of instrumental variables, which allows us now to control for the interaction between $Pull_{i,T}$ and lagged immigrant density on the right hand side of the second stage of the 2SLS procedure. We are now, effectively, relying only on differences in the diffusion process across metropolitan areas. The results are now larger, and so are the standard errors, which does not allow us to discard similar results to the ones found earlier.

Finally, in unreported work, we also modelled changes in immigrant shares using a logistic model (which bounds immigration shares between zero and one, and fits better the data in neighborhoods that are close to the data support boundaries), without substantial changes in the results. We present the simple linear interactions for ease of exposition.[13]

8 Rents

In Online Appendix Table 5 we show how the negative association between immigration and housing prices also holds for rents. We limit our sample to those metropolitan areas without rent control regulations. We use the initial number of renter households as weights in these regressions. The general associations with rents (columns 1 and 2) are weaker. However, their magnitude can be explained by the fact that rental units tend to be in areas denser with minority households and with lower housing quality. The interacted models posited in the paper, this time using data on rents (columns 3, 4, and 5), yield estimates that are surprisingly close to those in Table 2 of the paper. These results underline the robustness

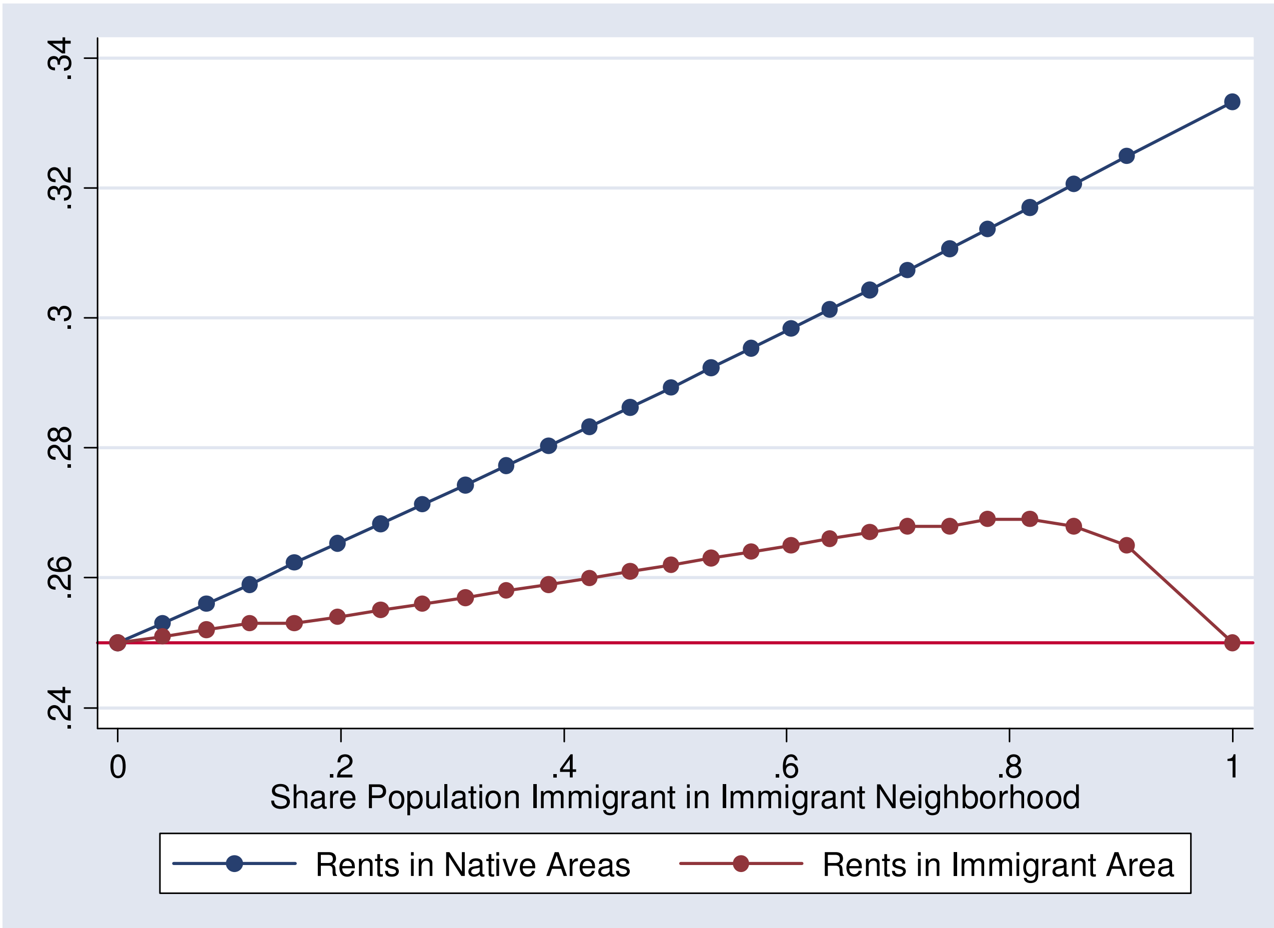
[13]Concretely, exclusion restrictions are easier to understand in the linear 2SLS version. More importantly, we wanted to be reassured that nonlinearities in the first stage did not play a major role in identification.

of the previous findings, and further reinforce a causal interpretation. While housing prices capitalize the impact of future events, rents reflect changes in the spot market demand for a neighborhood. Both sets of results are consistent with depressed contemporaneous and future market values *right after* an immigration shock.

9 SES Characteristics by Immigrant Group

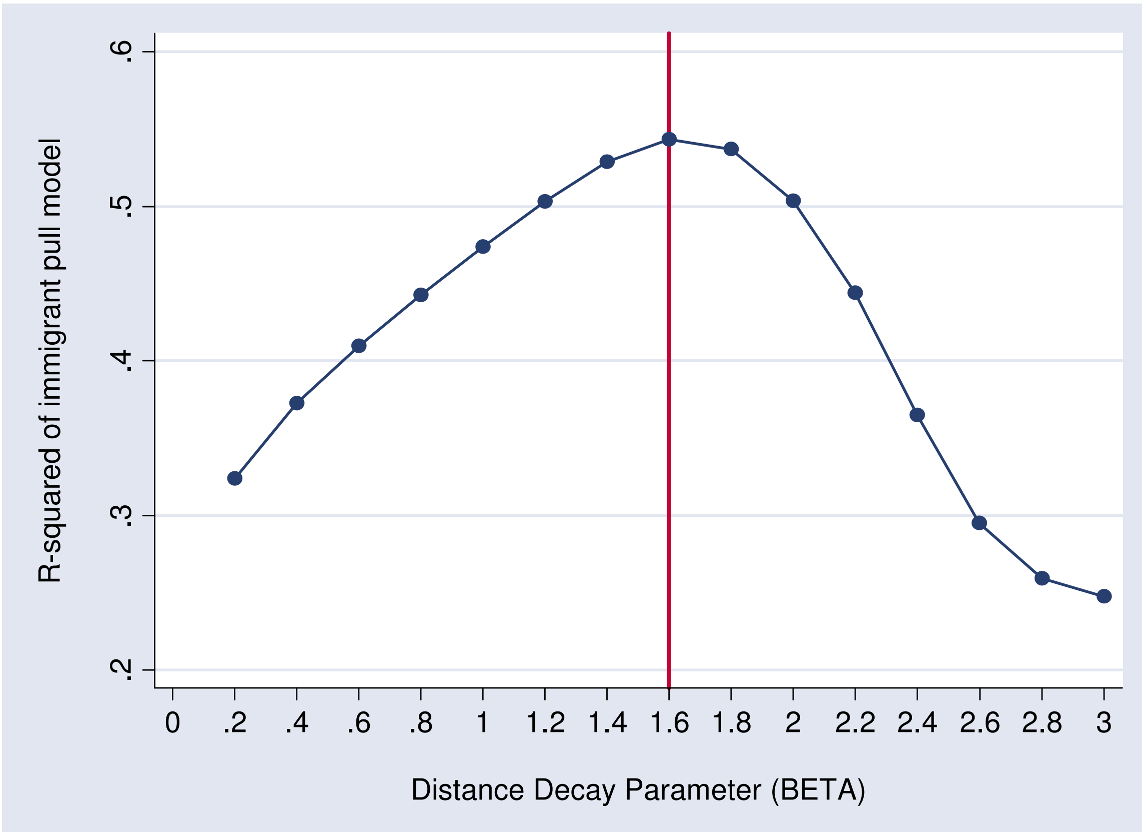
In the paper, we use immigrant characteristics by national group and state to infer the socio-demographic composition of local immigrant shocks. In Online Appendix Table 6, we provide the reader with the basic descriptives at the national level. We show data on the percentage of individuals without high school degree, bachelor's degree, and minority shares by national-geographic group. Mexican, Central-Americans and Dominicans in the US are amongst de groups with lower educational achievement. South Central Asians (from the Indian subcontinent), Africans, and Chinese in the US have the larger proportions of highly-educated immigrants (with BA/BS degrees). Europeans in the US are not particularly well-educated on average, but are, of course, mostly of European ancestry.

Online Appendix Figure 1
Immigrant Density and Housing Prices in a Simple Model

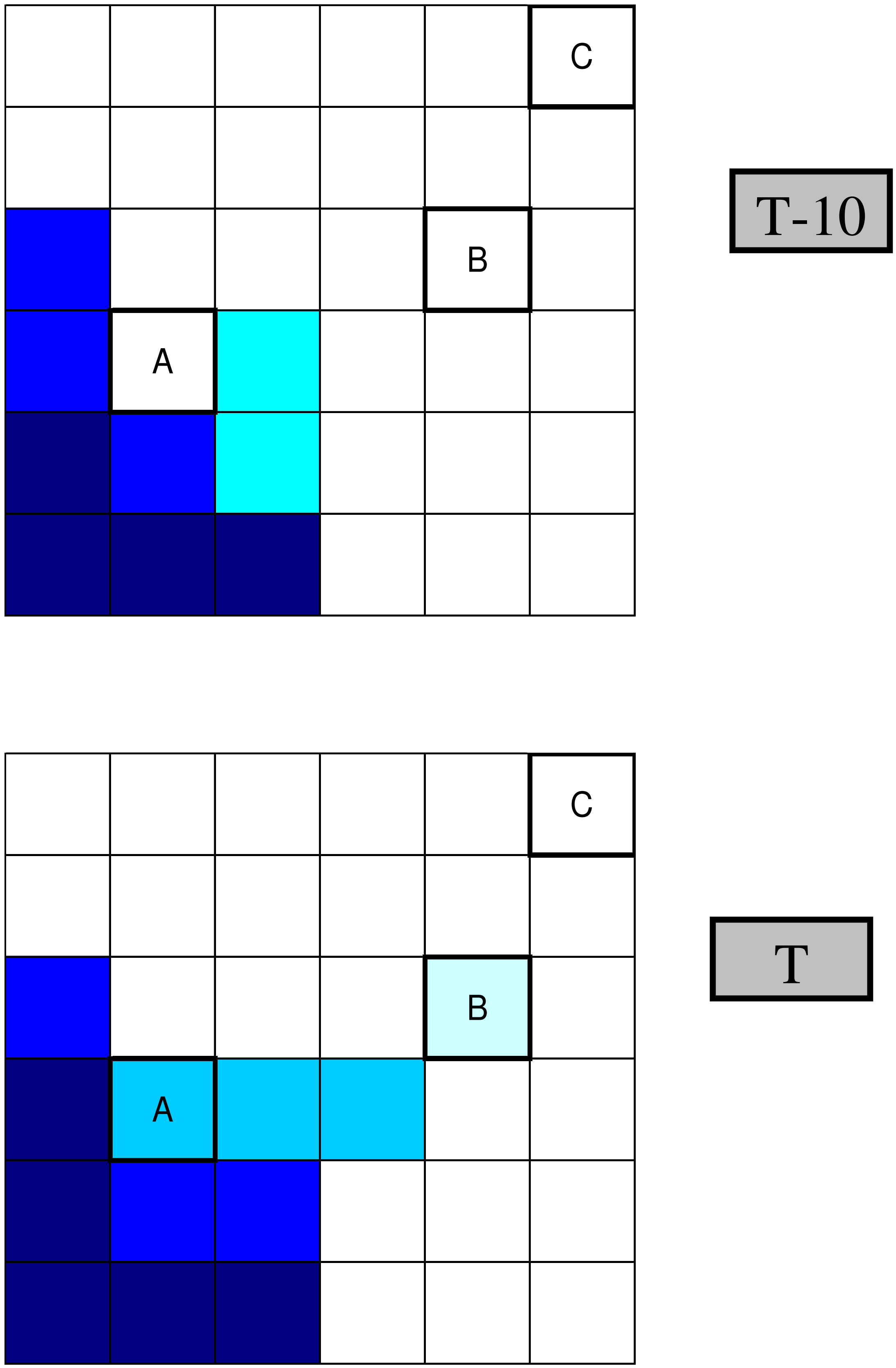


Online Appendix Figure 2

Spatial Correlation in Immigrant Settlement

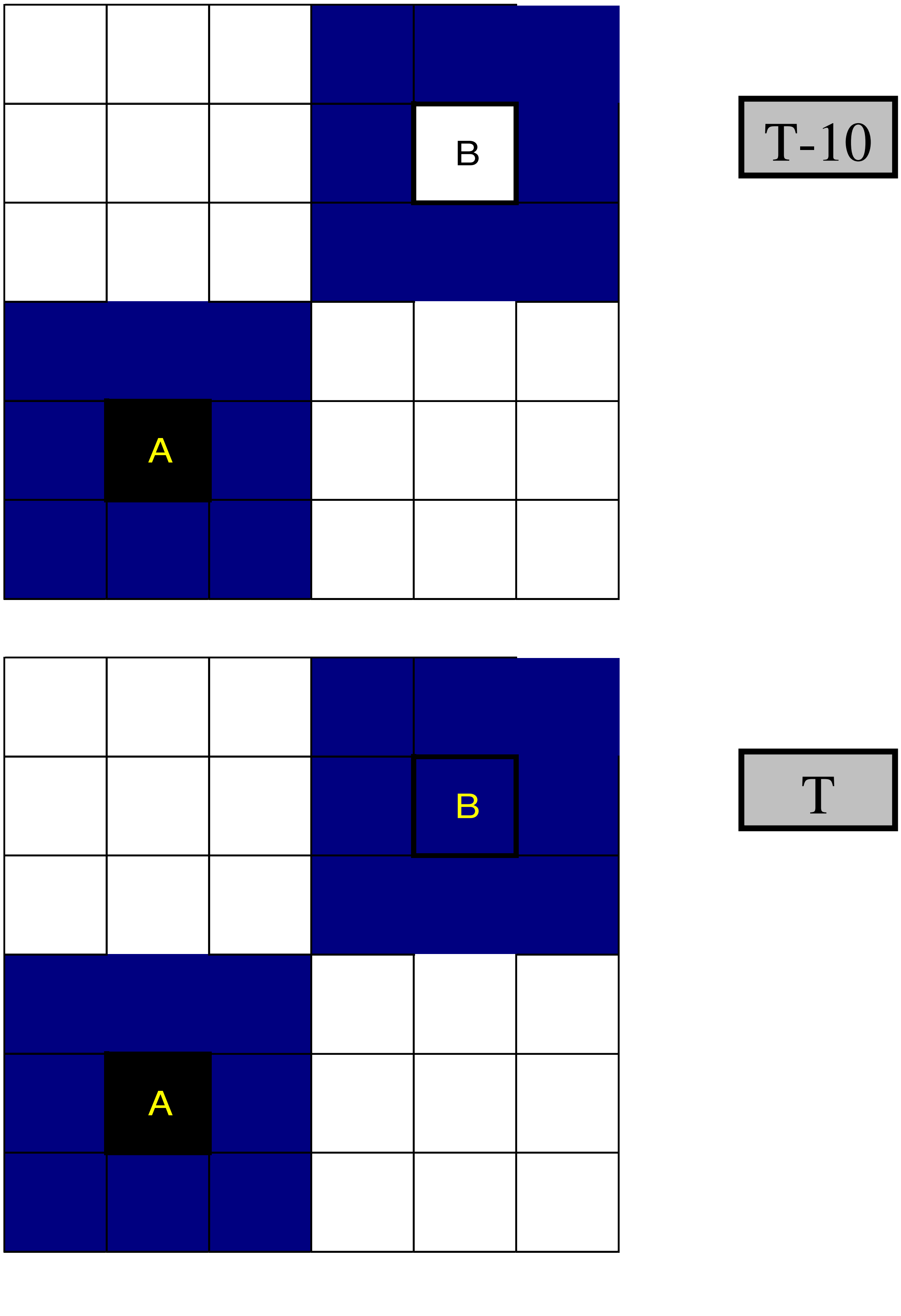


Online Appendix Figure 3
Diffusion of Immigrant Density (Illustration)



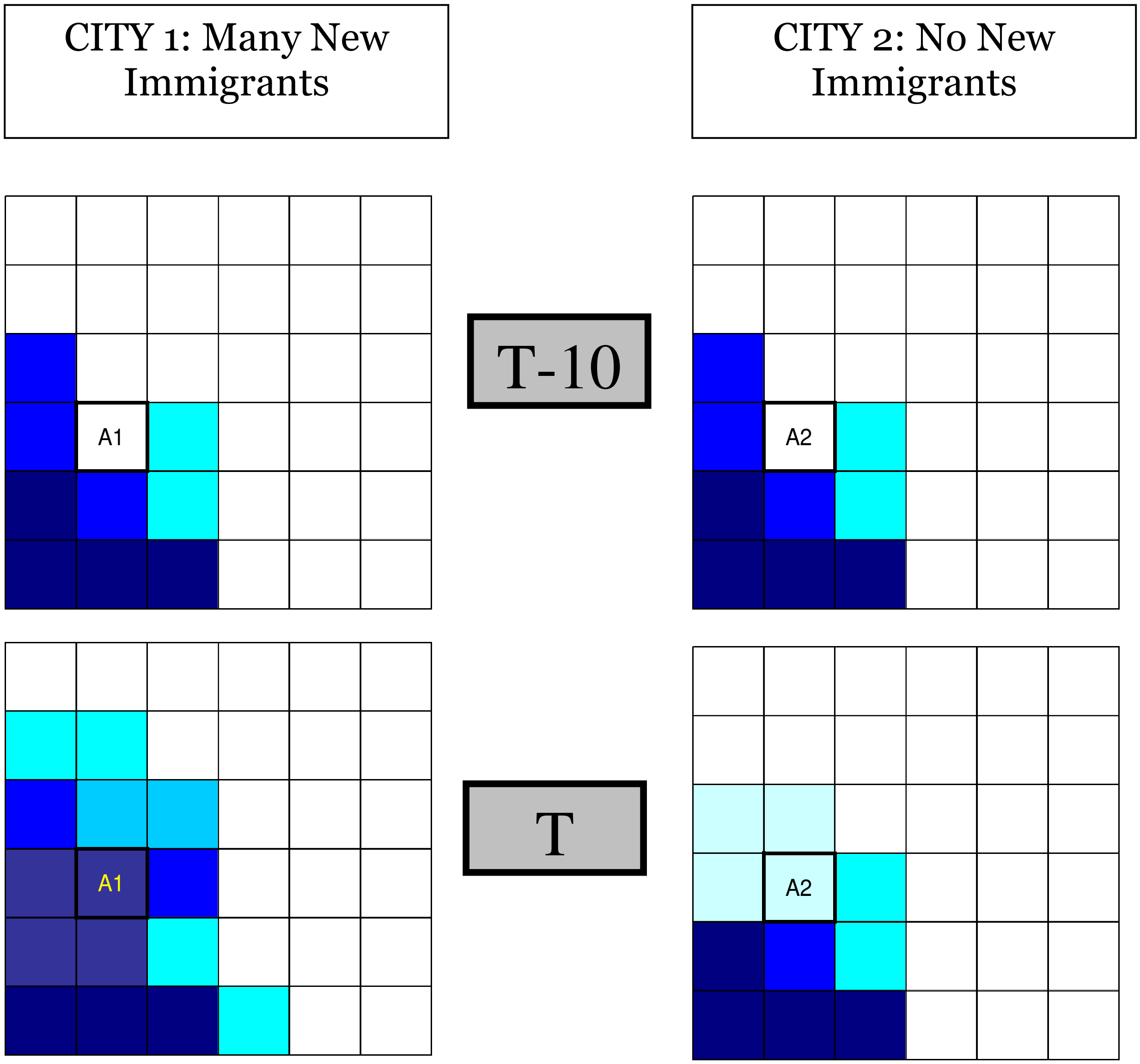
Online Appendix Figure 4

Diffusion of Immigrant Density. Similar neighbors, different initial immigrant densities (Illustration)

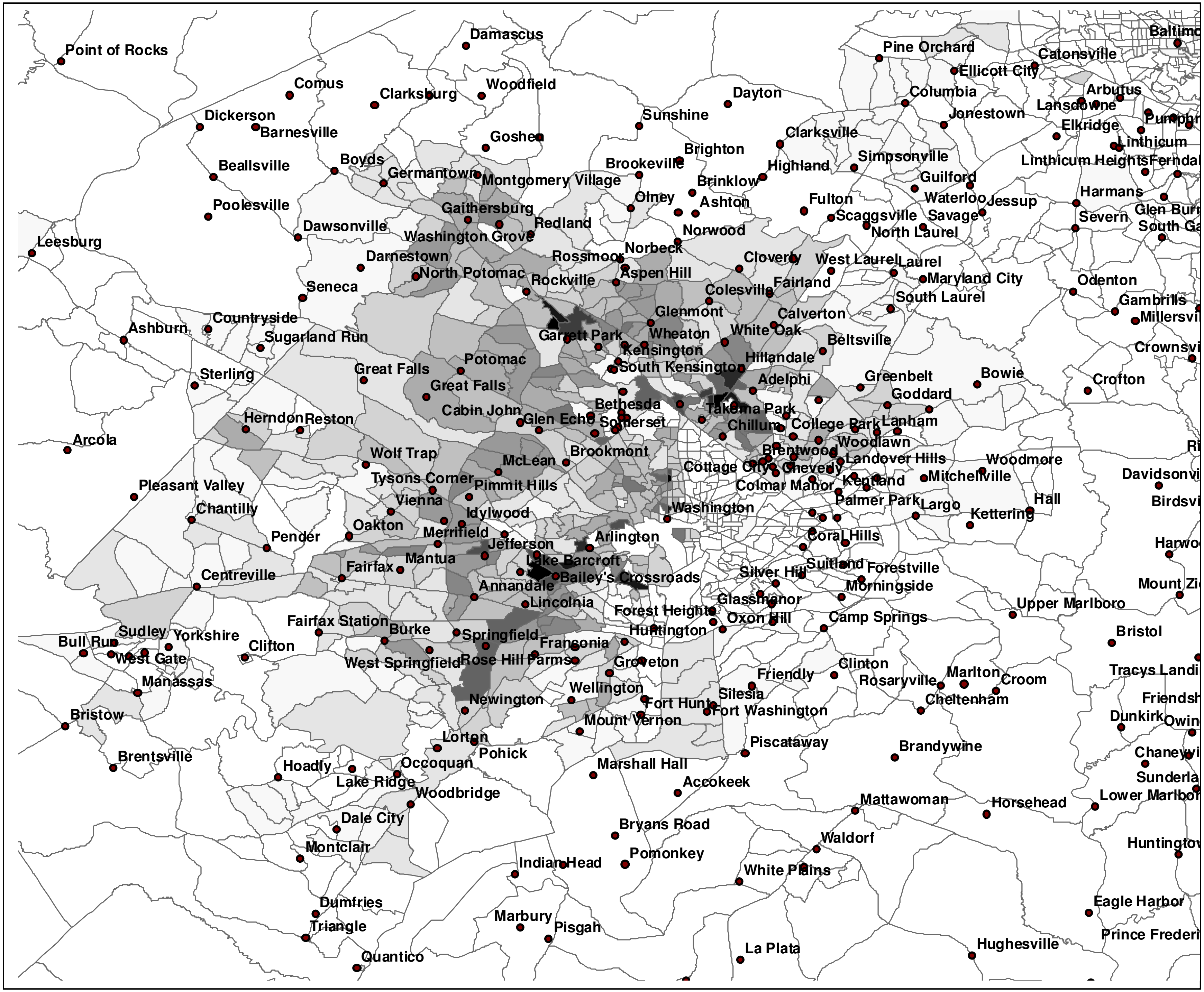


Online Appendix Figure 5

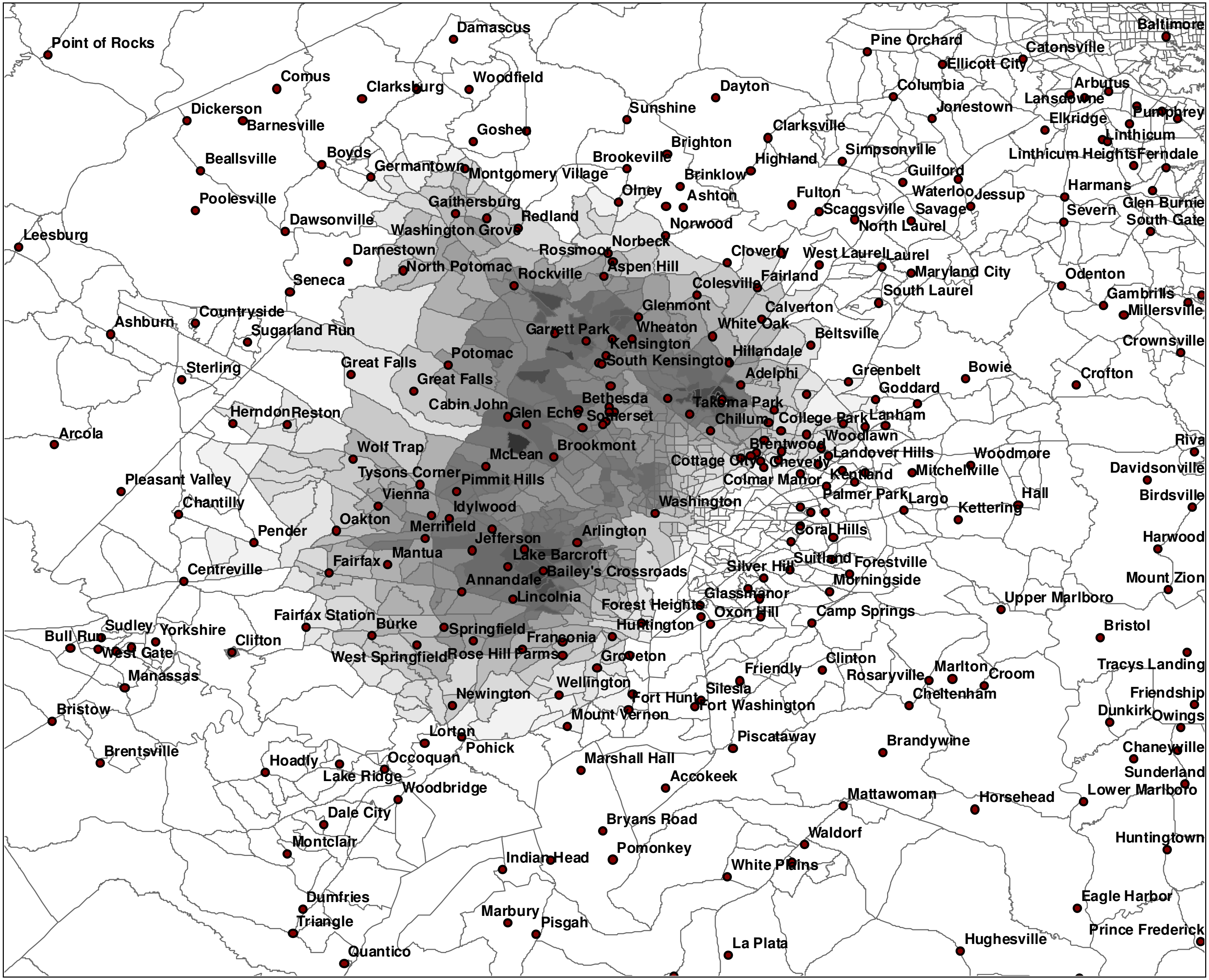
Diffusion of Immigrant Density. Similar neighbors, different immigration inflows at the MSA level (Illustration)



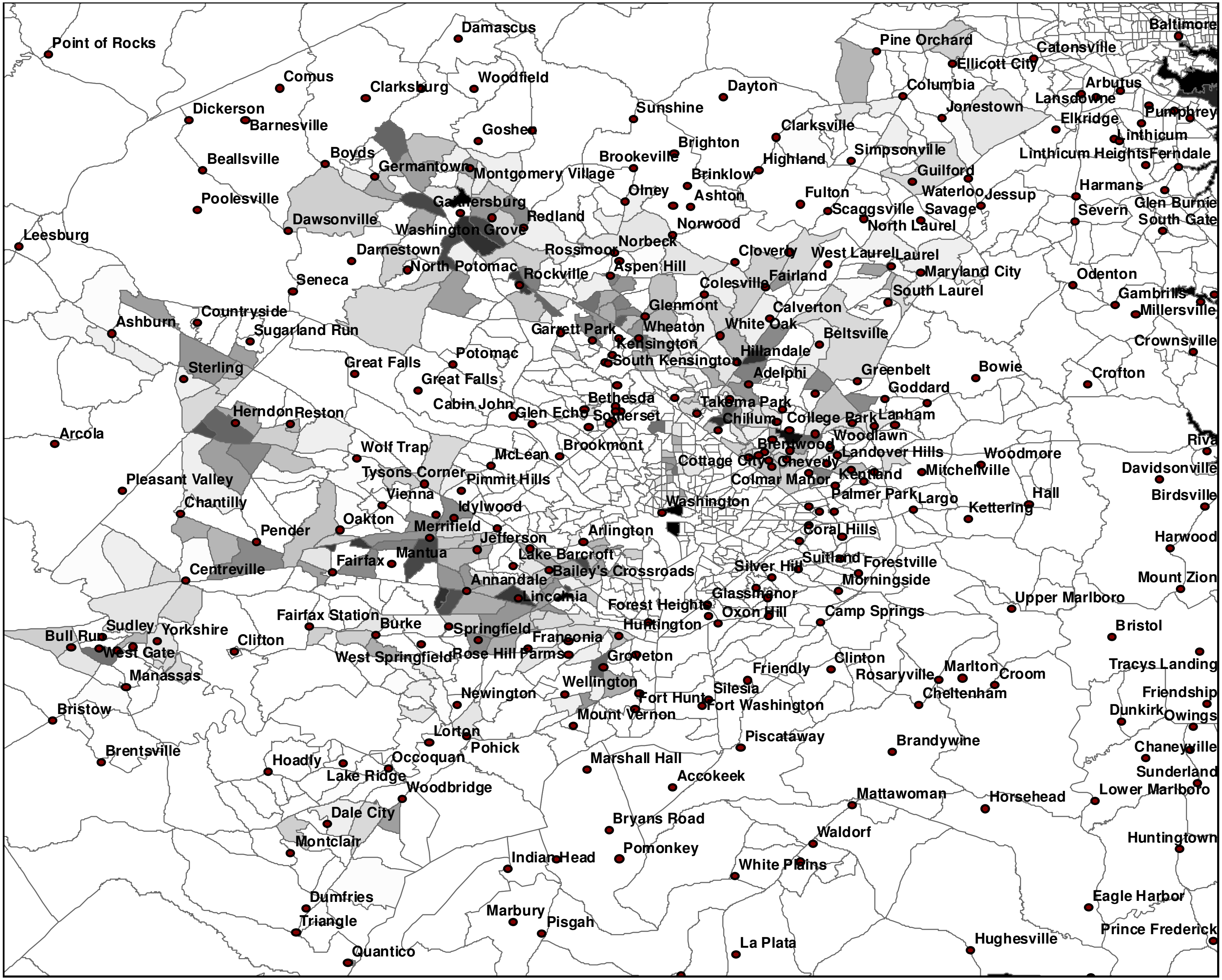
OA MAP 1: Immigrant Density in DC Area, 1990



OA MAP 2: Calculated Immigrant “Pull” in DC Area



OA MAP 3: Changes in Immigrant Density in DC Area, 1990-2000



Online Appendix TABLE 1

Immigrants and Housing Quality/Investments (AHS: 2001-2005)

	Total renovation costs		Open cracks wider than a dime		Large peeling paint areas		Water leak in basement		At least one room lacking electrical plugs		Neighborhood has neighborhood crime	
Foreign-Born Dummy (OLS/logit)	70.032	2.458	-0.079	-0.048	-0.119	-0.025	-0.572	-0.523	-0.092	-0.166	-0.066	-0.058
	(52.541)	(69.883)	(0.068)	(0.080)	(0.100)	(0.114)	(0.069)***	(0.080)***	(0.117)	(0.144)	(0.036)*	(0.043)
Foreign-Born Dummy (Logit FE)	210.02	123.355	-0.214	-0.162	0.618	0.635	-0.137	-0.147	-0.395	-0.381	-0.283	-0.27
	(259.796)	(259.423)	(0.221)	(0.227)	(0.396)	(0.413)	(0.237)	(0.241)	(0.389)	(0.404)	(0.140)**	(0.142)*
	Windows covered with metal bars		Windows broken		Holes/cracks or crumbling in		Roof has holes		Roof missing shingles/other roofing		Outside walls missing siding/bricks/etc.	
Foreign-Born Dummy (Logit)	1.298	1.374	-0.136	-0.189	-0.242	-0.235	0.011	-0.028	0.08	0.117	-0.083	-0.02
	(0.049)***	(0.059)***	(0.073)*	(0.088)**	(0.086)***	(0.105)**	(0.097)	(0.117)	(0.064)	(0.074)	(0.083)	(0.097)
Foreign-Born Dummy (Logit FE)	0.247	0.279	-0.534	-0.553	0.07	0.097	-0.071	0.009	0.245	0.278	0.051	0.078
	(0.268)	(0.280)	(0.325)	(0.346)	(0.339)	(0.347)	(0.401)	(0.416)	(0.236)	(0.239)	(0.299)	(0.304)
	Roof's surface sags or is uneven		Outside walls slope/lean/slant		Evidence of rodents		Garage or carport with unit		Holes in floor		Neighborhood has bad smells	
Foreign-Born Dummy (Logit)	-0.183	-0.22	-0.29	-0.251	-0.497	-0.475	0.14	0.16	-0.168	-0.149	-0.041	-0.018
	(0.091)**	(0.108)**	(0.125)**	(0.146)*	(0.034)***	(0.041)***	(0.028)***	(0.035)***	(0.150)	(0.176)	(0.057)	(0.066)
Foreign-Born Dummy (Logit FE)	-0.34	-0.33	0.659	0.878	0.064	0.062	-0.013	0.011	0.448	0.346	-0.078	-0.102
	(0.377)	(0.389)	(0.512)	(0.563)	(0.124)	(0.125)	(0.180)	(0.183)	(0.571)	(0.591)	(0.217)	(0.219)
Other Controls	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes

Notes:

Standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

The table displays the coefficient on a foreign-born dummy variable in separate regressions where a number of quality indexes are in the left-hand side. The foreign-born dummy takes value one if at least one member in the household is foreign born. The coefficients correspond to the parameter estimates in logit specifications, except in the case of renovation costs (OLS). The coefficients in the left of each set of regressions correspond to specifications where the only controls are the foreign-born dummy and year fixed effects. The coefficients in the right of each set of regressions correspond to specifications where we control for income, marital status, gender and age of reference person, year fixed effects, and a dummy for recent movers. The Logit and OLS regressions include data from the 2001, 2003, and 2005 waves of the AHS (upper rows). The fixed effect Logit specifications (bottom rows) consider only the subset of housing units that appear in both the 2001 and 2005 samples, and include housing unit fixed effects.

Online Appendix TABLE 2

OLS Robustness Tests

	∆Log Average Value: All MSA	∆Log Average Value: Native Neighborhoods	∆Log Average Value: Controls for Land Use in 1992	∆Log **Median** Value (1990-2000)	∆Log Average Value: Stratified Sample Average	∆Log Average Value: Controls for Past Immigrant Density
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
∆(Foreign Born/Population)	**-0.172**	**-0.276**	**-0.25**	**-0.278**	**-0.221**	**-0.245**
	(0.012)	(0.027)***	(0.015)***	(0.057)***	(0.021)***	(0.015)***
(Foreign Born/Population) at T-10						-0.028
						(0.012)**
MSA-Year Fixed Effects	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Change in Housing Quality	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Housing Quality at T-10	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Other variables in Table 1, Column 1	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Observations	96,152	17,364	34,492	21,681	34,833	34,833
R-squared	0.84	0.85	0.85	0.39	-	0.85

Notes:
Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
The table shows regressions where the change in the log of average (columns 1,3, and 4) or median (column 2)

Online Appendix TABLE 3

2SLS: First Stage

	(1)	(2)
Immigrant Gravity Pull	1.909	2.699
	(0.343)***	(0.993)***
Share Foreign Born at T-10		0.292
		(0.010)***
Gravity Pull • Share Foreign Born at T-10		-17.243
		(0.735)***
Gravity Pull • (MSA Immigrants/Initial Population)		31.877
		(4.373)***
MSA-Year Fixed Effects	*yes*	*yes*
Other Variables in Table 1, Column 2	*yes*	*yes*
Observations	34,833	34,833
R-squared	0.27	0.32

Notes:

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

The table shows the first-stage regressions of the 2SLS approach in Table 4.

Online Appendix TABLE 4

IV: Robustness Tests

	ΔLog Average Value			
	Non-lineary Controls for Immigrant Gravity **(1)**	*Controls for Changes in Immigrant Concentration of Neighboring Tracts* **(2)**	*Controls for Lagged Neighbor Price Evolution* **(3)**	*Only Variation from MSA Immigration Levels* **(4)**
Δ(Foreign-Born/Population)	**-0.210**	**-0.208**	**-0.183**	**-0.364**
	(.063)***	(0.056)***	(0.052)***	(0.198)*
Controls for Gravity Pull	yes	yes	yes	yes
Controls for (Gravity Pull×Share Foreign Born at T-10)	no	no	no	yes
Other variables in Table 1, Column 2	*yes*	*yes*	yes	*yes*
MSA-Year Fixed Effects	yes	yes	*yes*	*yes*
Instruments	*Gravity Pull×MSA Immigration, Gravity Pull×Share Foreign Born at T-10*	*Gravity Pull×MSA Immigration, Gravity Pull×Share Foreign Born at T-10*	*Gravity Pull×MSA Immigration, Gravity Pull×Share Foreign Born at T-10*	*Gravity Pull×MSA Immigration, Gravity Pull×Share Foreign Born at T-10×MSA Immigration*
F-test of excluded variables	142.14		423.497	44.27
Hansen Overidentification Test (p-values)	0.70		0.43	0.00
N	34,833		34,833	34,833

Notes:
Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%
+ As defined in text and divided by 1,000,0000
The table shows regressions where the change in the log of average housing prices between consecutive decennial censuses by census tract is the left-hand·

Online Appendix TABLE 5

Immigrant Density and Rents Within the City

	ΔLog Rents				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Δ(Foreign Born/Population)	-0.284	-0.079	0.035	-0.011	0.042
	(0.016)***	(0.016)***	(0.04)	(0.02)	(0.04)
Δ(Foreign Population/Population) • Share Non-Hispanic white at T-10			-0.196		-0.102
			(0.059)***		(0.062)*
Δ(Foreign Population/Population) • House Value Quartile at T-10				-0.058	-0.051
				(0.013)***	(0.014)***
MSA-Year Fixed Effects	*yes*	*yes*	*yes*	*yes*	*yes*
Change in Housing Quality	*no*	*yes*	*yes*	*yes*	*yes*
Housing Quality at T-10	*no*	*yes*	*yes*	*yes*	*yes*
Other variables in Table 1	*no*	*yes*	*yes*	*yes*[╨]	*yes*[╨]
Observations	21,295	21,282	21,282	20,694	20,694
R-squared	0.65	0.73	0.73	0.74	0.74

Notes:

Robust standard errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

[╨] In equations 2 and 3, we substitute log of income at T-10 by log of housing values at T-10. The

The table shows regressions where the change in the log of average housing rents between consecutive

Online Appendix TABLE 6

Main SES Characteristics by World Region

World Region	Percent Dropout	Percent with Bachelor's Degree	Percent Minority (Hispanic or Nonwhite)
Africa	7.18	45.95	68.83
South Central Asia	10.57	60.70	94.32
Philippines	12.08	44.68	96.43
Middle East	13.70	43.05	11.68
South America	18.77	23.01	88.54
Caribbean	21.09	15.87	97.04
China	21.54	44.99	97.91
East Asia	21.73	28.33	94.23
Europe	22.33	24.96	7.69
Other	31.56	22.92	56.79
Cuba	33.39	17.51	98.37
Dominican Republic	45.56	8.98	99.01
Central America	46.31	9.32	96.34
Mexico	65.46	4.20	99.18