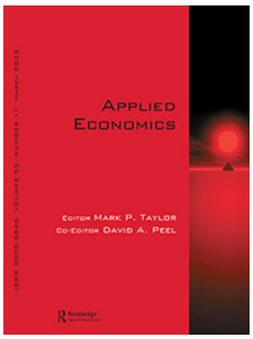

Applied Economics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raec20

Language and segregation: evidence from housing markets in the United States

Tiantian Dai, Shenyi Jiang, Tao Jin & Buyuan Yang

To cite this article: Tiantian Dai, Shenyi Jiang, Tao Jin & Buyuan Yang (2023) Language and segregation: evidence from housing markets in the United States, Applied Economics, 55:11, 1157-1183, DOI: 10.1080/00036846.2022.2096866

To link to this article: https://doi.org/10.1080/00036846.2022.2096866

Published online: 19 Jul 2022.

APPLIED ECONOMICS
2023, VOL. 55, NO. 11, 1157–1183
https://doi.org/10.1080/00036846.2022.2096866

Routledge
Taylor & Francis Group

Language and segregation: evidence from housing markets in the United States

Tiantian Dai[a], Shenyi Jiang[a], Tao Jin[b] and Buyuan Yang[c]

[a]China Economics and Management Academy, Central University of Finance and Economics, Beijing, China; [b]PBC School of Finance, & Hang Lung Center for Real Estate, Tsinghua University, Beijing, China; [c]Institute of Economics, School of Social Sciences, Tsinghua University, Beijing, China

ABSTRACT

It has been suggested that residential segregation has negative social and economic impacts. However, psychological and cultural benefits of such clustering have been established by research. We undertake a comprehensive analysis by estimating residential choices of minorities and identifying their willingness to pay for housing attributes and community characteristics. Recognizing that the language spoken at home adds to daily convenience, a sense of belonging, and a feeling of closeness, we widen the scope of the study to include four minority groups, defined by the languages spoken at home. We find that preferences for integrating into white, English-speaking societies are quite heterogeneous. The willingness to pay to live in a community with more own language speakers is inversely associated with the willingness to pay to live in a community with more English-speaking whites. Furthermore, any two minority groups share the same reciprocal attitudes towards each other. To assimilate more into the mainstream, a minority group might choose to reside in a community with fewer people who speak their language, but they would not necessarily be willing to lower the percentage of other minority groups.

KEYWORDS

Hedonic price function; non-parametric estimation; residential segregation

JEL CLASSIFICATION

C5; H0; J7; R0; R2

I. Introduction

Individuals may be segregated by ethnicity, income, occupation or preference for local public goods such as access to public transit (Gradín 2020; Moulin, Flacher, and Harari-Kermadec 2016; Alonso-Villar, Gradin, and Del Río 2013; Conejeros and Vargas 2012). Residential segregation is a key dimension for measuring the pattern of segregation. It is commonly observed and extensively studied in several countries with large-scale immigration (Bayer, McMillan, and Rueben 2004a; Bayer, Ferreira, and McMillan 2007; Bayer and McMillan 2012; Kuminoff, Smith, and Timmins 2013).

Residential segregation can cause adverse social and economic effects. Living in highly segregated areas limits job opportunities and access to high-quality local public goods, as well as the means for sharing the spillover benefits from more skilled workers (Kain 1968; Borjas 1995; Wilson 2012). In addition to the direct social and economic effects, people have developed a distaste for segregation as it contradicts the ideology of a free and equal society (Zhang and Zheng 2015). Thus, the experience in the United States of 'ghettoization' symbolizes persistent Black minority segregation as a failure of efforts towards social and economic integration and equality of opportunity in housing and employment (Bolt, Phillips, and Kempen 2010).

In addition to the negative impacts, the psychological and cultural benefits of ethnic clustering have been well documented (Peach 1996; Phillips 2007); however, it is rarely measured by researchers due to the difficulty of doing so. Extended social relations, cultural support, the feeling of belonging, and community infrastructure oriented towards ethnic goods and services may result in a sense of well-being and security (Phillips, Davis, and Ratcliffe 2007). Gilroy (2013) shows that because of the feelings of community and belonging, the 'ghetto' serves as a place for possible resistance rather than feelings of victimization and exclusion. In the United Kingdom in recent years, minority ethnic groups have increasingly chosen 'self-segregation' and withdrawal from British society to live 'parallel lives' (Phillips 2006).

CONTACT Shenyi Jiang syijiang@gmail.com China Economics and Management Academy, Central University of Finance and Economics, Beijing, China; Buyuan Yang yby19@mails.tsinghua.edu.cn Institute of Economics, School of Social Sciences, Tsinghua University, Beijing, China

A comprehensive welfare analysis of residential segregation must take both the positive and negative effects into account. This paper contributes to the literature by estimating the willingness of minorities to pay for segregation, thereby capturing the overall effects. Households derive utility from housing attributes and community characteristics. Neighbourhoods vary in future opportunities and levels of segregation. All else equal, households face a trade-off when choosing where to live. If a household is willing to give up cultural and social support to live in a less segregated community, then segregation must be undesirable; otherwise, if the opposite is true, segregation must be a preferred attribute.

This study differs from others in two distinct aspects. First, it is one of the few studies to single out the role of the language spoken at home in shaping the landscape of residential segregation. Minorities may be forced to speak English at work; however, they may prefer speaking their mother languages at home or in the community. Speaking the same language helps to provide daily convenience, a sense of belonging, and a feeling of closeness. Such commonness intensifies the positive side of residential segregation and enhances the likelihood of clustering. Previous studies (Massey and Denton 1987, 1989; Bayer, McMillan, and Rueben 2004b; Wong 2013) have stressed the effects of race. Without denying the importance of race, we consider it as a broader and more multidimensional concept compared with the language spoken at home. Depending on their origin, people who identify as a single race might speak multiple languages and accommodate a wider range of diversity compared with a single mother language group. They may vary considerably in habits, customs, religions, or even values, all of which are reflected in their residential choices. Furthermore, some racial minorities, mainly the second or third generation of immigrants, speak only English and they differ in many ways from the first generation of immigrants. These two scenarios call for a re-investigation of residential choices using a different grouping method. Therefore, we focus on spoken language in this study.

The second major feature of our paper is the way we measure segregation. We make a distinction between two types of segregation: spatial segregation (or residential segregation) and preference segregation. A group may be physically segregated if they don't reside near English-speaking white community (mainstream). Still, they are not necessarily be 'preference-segregated' unless they were willing to pay more to live with the group speaking the same language. It's worth noting that spatial segregation is an optimal choice made by households, jointly determined by preference separation (more of preference) and other constraints such as financial constraints. Likewise, the same distinction applies to the integration. Most the previous studies have relied on the dissimilarity index to measure segregation (Massey and Denton 1987, 1989; Yalonetzky 2012; Allen et al. 2015), which is the spatial segregation that has been defined. Moreover, without adjusting for variations in the diversity of population distributions, this type of measurement often provides an average composition of the local environment (Reardon and O'Sullivan 2004). Instead of using the dissimilarity index to measure spatial segregation, we focus more on preference segregation.[1] We utilize a structural estimation and infer the degree of segregation from the willingness to pay for a higher penetration rate of different language speakers. Compared to the dissimilarity index, the preference segregation is an indirect approach; however, it reveals minority groups' subjective feelings about segregation more accurately.

We apply the hedonic model to estimate each minority language group's willingness to pay for housing attributes and community characteristics and evaluate the level of preference segregation. The hedonic model effectively characterizes individual preferences, especially in the housing market. This method has been adopted by several studies (Bartik 1987; Epple 1987; Bajari and Benkard 2005; Koster, Van Ommeren, and Rietveld 2016; Bayer et al. 2016). Additionally, the hedonic method is applied to nonparametric estimation of partial equilibrium models (Sieg et al. 2004; Epple, Peress, and Sieg 2010; Kuminoff, Smith, and Timmins 2013).

[1]Unless otherwise specified, the segregation indicated in the main text is referred to the preference segregation.

The methodology here is closely related to that of Bajari and Benkard (2005), who develop a semi-parametric hedonic method to estimate households' heterogeneous housing demand, thereby allowing preferences to be a flexible function of observed demographics and idiosyncratic taste shocks for housing characteristics. Furthermore, to meet our research objectives, we extend the methodology to more than two minority groups.

In this study, the minority groups of focus include Hispanics (Spanish speakers), Chinese (Chinese speakers), Indians (Hindi and related language speakers), and English-speaking groups.[2] Albeit drawing much less academic attention, these minorities represent an increasingly high percentage of the US population. The patterns of social and spatial incorporation are likely to be differentiated over time and vary among places (Phillips 2007; De la Roca, Ellen, and O'Regan 2014). It would be worth investigating whether the pattern of segregation and attribution of these minority groups are systematically different from each other. The flexibility of the estimation strategy facilities the study of minority groups' segregation from or integration into the mainstream (English-speaking white communities) and preferences towards each other. To the best of our knowledge, this is the first paper to study the current status of minority (other than Black) integration by analysing individuals' residential choices.

In line with Bajari and Kahn (2005), for the purpose of direct comparison, we sample the same three representative metropolitan cities: Chicago, Dallas, and Atlanta. Except for Chinese speakers, we find strong evidence in favour of urbanization among the minority groups, in the sense that they neither value more rooms or single detached houses, nor do they care much about housing ownership or the age of buildings. The demand for higher-educated peers varies across cities and groups. Compared with previous literature, we contribute to the existing segregation literature on other minority groups. Furthermore, we estimate the willingness to pay for a higher penetration rate of own or other language speakers, which sheds light on the segregation situation for each minority group. Based on the willingness to pay for a higher percentage of English-speaking white group, Chicago is the least integrated city and Dallas is at the other extreme, consistent with previous studies. In addition, we find that for each minority language group, the willingness to pay for more own language speakers and the willingness to pay for more English-speaking white are largely negatively correlated. Any two minority groups share the same reciprocal attitudes towards each other. A corollary inference can be made based on the patterns discovered. To integrate more into the mainstream by increasing the share of English-speaking white, a minority group would choose to have fewer own language speakers, but may not necessarily be willing to lower the percentage of other minority groups. To be away from the mainstream, a minority group would like to substitute with more own language speakers.

The rest of the paper is organized as follows. Section II introduces the hedonic model to estimate households' willingness to pay. Section III presents the data and the baseline model. Section IV estimates the willingness to pay for housing attributes and community characteristics and the demand for integration and segregation. Section V provides further discussion, and section VI concludes.

II. Methodology

Basic model

We first introduce the basic model and estimation methodology. Following Bajari and Kahn (2005), we assume that a residence is composed of three sets of attributes: physical attributes, community attributes, and attributes only observed by households. Let $i = 1, 2, \ldots, I$ denote households and $j = 1, 2, \ldots, J$ denote housing units. The physical attributes include the number of rooms ($Rooms_j$), the age of the residence ($Builtage_j$, years), single detached or not ($Single_j$), and the property ownership

[2]Ideally, it would be desirable to have English speakers as one group in the study. However, the issue is that there is a vast discrepancy inside the English-speaking group. Even if we proceeded by treating English speakers as a single group, the estimated willingness to pay for is difficult to interpret.

(Own_j, $1 = own$, $0 = rent$). The community attributes include the percentage of college educated households in the community ($mCollege_j$), and the percentages of minority language speakers ($mSpanish_j$, $mHindi_j$, $mChinese_j$). Moreover, we assume the prior percentage of English speakers. Considering the profound heterogeneity between the white and English-speaking non-white groups, we assume the percentage of English-speaking white ($mEnglishwhite_j$) and the percentage of English-speaking non-white groups ($mEnglishnonwhite_j$) simultaneously. Let ξ_j denote the unobserved attribute of housing unit j, and c is the consumption of a composite commodity. We assume that household i aims to maximize the following utility function:

$$\begin{aligned} u_{i,j} = {} & \beta_{i,1} log(Rooms_j) + \beta_{i,2} log(Builtage_j) \\ & + \beta_{i,3} Single_j + \beta_{i,4} Own_j + \beta_{i,5} log(mSpanish_j) \\ & + \beta_{i,6} log(mHindi_j) + \beta_{i,7} log(mChinese_j) \\ & + \beta_{i,8} log(mEnglishwhite_j) \\ & + \beta_{i,9} log(mEnglishnonwhite_j) \\ & + \beta_{i,10} log(mCollege_j) + \beta_{i,11} log(\xi_j) + c. \end{aligned} \tag{1}$$

To recover individual-specific taste coefficients, we further specify an linear model of household demographics in Equation (2), shown as followed:

$$\beta_{i,k} = \theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} + \eta_{i,k}, \tag{2}$$

where $k = 1, 2, 5, \ldots, 10$ and $\beta_{i,k}$ denotes the household-specific preference parameter. When $k = 3$ and $k = 4$, $\beta_{i,k}$ denotes the coefficients of the dichotomous variables, which are estimated by probit models. The estimation strategy is stated in the following section. $\eta_{i,k}$ is the household $i's$ specific taste shock for preference parameter $\beta_{i,k}$. d_i is a set of demographic characteristics of the household head: age (*Age*, 18–60 years old), marital status (*Married*, 1 = married, 0 = not), gender (*Male*, 1 = male, 0 = female), education (*College*, 1 = college graduate, 0 = not), language spoken at home (*Spanish*, *Hindi and related*, *Chinese*, all dichotomous variables), the household's total annual income (*Hhincome1000*, measured in thousands of 2018 dollars), family size (*Famsize*), and an interaction term between non-white and English language (*Englishnonwhite*) considering the distinctive patterns of behaviour of English-speaking white and non-white groups. Language could have various effects across different races.

We assume that continuous preference parameters can be expressed as linear functions of demographics. The household-specific taste shock $\eta_{i,k}$ is orthogonal to the preference parameter with $E(\eta_i|d_i) = 0$. Once the continuous preference parameters are recovered, individual-specific taste coefficients $\theta_{k,s}$ can be estimated by running linear regressions.

Estimation procedure

In market m, a household i's maximization problem is

$$\max_j \; u_i(\boldsymbol{x}_j, \xi_j, y_i - \boldsymbol{P}_m(\boldsymbol{x}_j, \xi_j)), \tag{3}$$

where, household i choose housing units j to maximize her utility in market m, and the optimal choice is denoted by j^*. $\boldsymbol{x}_j$ is a matrix including all observed product characteristics. y_i denotes household i's total annual income, and $\boldsymbol{P}(\boldsymbol{x}_j, \xi_j)$ is the housing price function. The subscript $m = 1, 2, \ldots, M$ denotes a metropolitan area. Each metropolitan city is assumed to be isolated from other areas; therefore, we conduct the estimations separately and drop the market subscript m in the following sections.

Estimating continuous variables

To maximize utility, we take the first-order conditions with respect to the continuous variables and obtain

$$\beta_{i,k} = x_{j^*,k} \frac{\partial \boldsymbol{P}(\boldsymbol{x}_{j^*}, \xi_{j^*})}{\partial x_{j,k}} \tag{4}$$

To recover the preference parameters, we apply the hedonic approach proposed by Bajari and Kahn (2005).[3] We assume that the local non-parametric hedonic price function has the following log-linear form for each market:

$$\begin{aligned} p_j = {} & \alpha_{0,j^*} + \alpha_{1,j^*} log(Rooms_j) + \alpha_{2,j^*} log(Builtage_j) \\ & + \alpha_{3,j^*} Single_j + \alpha_{4,j^*} Own_j \\ & + \alpha_{5,j^*} log(mSpanish_j) + \alpha_{6,j^*} log(mHindi_j) \\ & + \alpha_{7,j^*} log(mChinese_j) \\ & + \alpha_{8,j^*} log(mEnglishwhite_j) \\ & + \alpha_{9,j^*} log(mEnglishnonwhite_j) \\ & + \alpha_{10,j^*} log(mCollege_j) + \xi_j, \end{aligned} \tag{5}$$

where, $\alpha_{k,j^*}(k = 0, 1, \ldots, 10)$ is the hedonic coefficient for housing unit j^*. The implicit price of j^* depends on the physical attributes, community attributes, and the unobserved attribute. Each α_{k,j^*} estimates how much an independent variable contributes to the price of the housing unit j^*. This estimation is executed at the city level. ξ_j denotes the unobserved attribute of each housing unit.

As suggested by Fan and Gijbels (2018), the hedonic coefficients can be estimated by weighted least squares (WLS) with variable weights for each residence. Therefore, for each j^*, we estimate the corresponding hedonic coefficients as follows:

$$\alpha_{j^*} = \underset{\alpha}{\operatorname{argmax}} (\boldsymbol{P} - \boldsymbol{X}\alpha)' \boldsymbol{W} (\boldsymbol{P} - \boldsymbol{X}\alpha), \tag{6}$$

where

$$\boldsymbol{P} = [p_j] : (J \times 1),$$

$$\boldsymbol{X} = [\boldsymbol{x}_j] : (J \times 10),$$

$$\boldsymbol{W} = diag\{K_h(\boldsymbol{x}_j - \boldsymbol{x}_{j^*})\} : (J \times J).$$

J is the number of housing units in each metropolitan area m. $\mathbf{W}$ is a $J \times J$ diagonal matrix of kernel weights. Each element on the diagonal is a normal kernel function with bandwidth equal to 3 (that is, $h = 3$):

$$K(\boldsymbol{x}_j - \boldsymbol{x}_{j^*}) = \prod_{k=0}^{10} N((\boldsymbol{x}_j(k) - \boldsymbol{x}_{j^*}(k))/\hat{\sigma}_k^2),$$

$$K_h(\boldsymbol{x}_j - \boldsymbol{x}_{j^*}) = K((\boldsymbol{x}_j - \boldsymbol{x}_{j^*})/h)/h,$$

where N is the probability density function (PDF) of the standard normal distribution of each characteristic k. K is a function equal to the product of all the standard normal PDFs. $\hat{\sigma}_k^2$ is the standard deviation of characteristic k.

Following the hedonic literature, we assume that the unobserved characteristic ξ_{j^*} is independent of the observable variables. After conducting the WLS and hedonic regressions, ξ_{j^*} and the household-specific preference parameters for the continuous variables can be calculated as $p_{j^*} - \hat{p}_{j^*}$ and

$$\hat{\beta}_{i,k} = \hat{\alpha}_{k,j^*}. \tag{7}$$

Next, we estimate the individual-specific taste coefficients $\hat{\theta}_{k,s}$ by running the following OLS regression:

$$\hat{\beta}_{i,k} = \theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} + \eta_{i,k}, \tag{8}$$

where the residual $\eta_{i,k}$ is the household-specific taste shock. We model the joint distribution of tastes and demographic characteristics using a linear model. The interaction term between white and English speakers is not included, as it is controlled by one of the dummies.

Dichotomous variables

This section describes how to identify the coefficients $\beta_{i,3}$ and $\beta_{i,4}$ of the two dichotomous variables, namely, a single detached dwelling and an ownership. These are dummy variables estimated using probit models. We assume that the household-specific taste shock $\eta_{i,k}$ follows a normal distribution. The probability of household i choosing to own a single detached house is

$$N\left(\theta_{0,k} + \sum_s \theta_{k,s} d_{i,s} - \frac{\Delta p}{\Delta \text{Single}}; 0, \sigma\right), \tag{9}$$

where N is the CDF of the normal distribution with mean zero and standard deviation σ.

A household would choose to live in a single detached house only if the marginal benefit is at least as much as the implicit marginal cost. Here

[3] In Bajari and Kahn's paper, they try different methods including Hedonic methods-local linear, high-order polynomials and kernel regression. The Hedonic estimation outperforms the other two. See Bajari and Benkard (2005), Bartik (1987) and Epple (1987) for further references.

the marginal cost $\frac{\Delta p}{\Delta Single}$ is in fact the estimated parameter $\alpha_{3,j^*(i)}$ from Equation (5). Thus, in the likelihood function of the dummy variable *single detached house*, we replace $\frac{\Delta p_m}{\Delta Single}$ by $\alpha_{3,j^*(i)}$. Moreover, we confine the coefficient of the implicit price to be -1. The likelihood function of the *single-detached house* dummy variable is as follows:

$$L(\theta;0,\sigma) = \prod_{i=1}^{I}\left[1-N\left(\theta_{0,k}+\sum_s \theta_{k,s}d_{i,s}-\alpha_{3,j^*(i)};0,\sigma\right)\right]^{1-Single_{j^*(i)}} \times\left[N\left(\theta_{0,k}+\sum_s \theta_{k,s}d_{i,s}-\alpha_{3,j^*(i)};0,\sigma\right)\right]^{Single_{j^*(i)}}. \quad (10)$$

The likelihood function of the *own* dummy is defined in the same way.

III. Data and empirical motivation

Data

The data on household head demographics and housing unit characteristics were drawn from the American Community Survey 2018 (ACS2018), a data set downloaded from the US database known as the Public Use Microdata Series (IPUMS). ACS2018 specifically surveys the languages households speak at home. We use this question to infer the preference of language used in daily communication, assisting in determining the mother language for each household head.

We select Chicago, Atlanta, and Dallas as the targeted metropolitan cities in this analysis. There are three reasons to confine the study to these cities. First, we intentionally avoid the areas with extremely high housing prices. Expensive residence requires strong abilities to pay, imposing a hard constraint on purchasing decisions. Those who choose to reside in expensive areas are among the top percentiles of income distribution and might systematically differ themselves from the rest of the population. Thus we exclude expensive areas, such as New York City and Boston, etc. Second, We avoid cities with potential impacts of occupational segregation, such as Seattle and San Francisco, etc. Third, we select the same three cities with Bajari and Kahn (2005) in order to be able to compare our results with theirs. The cross-city analysis also facilitates conducting a comparison study. Based on the dissimilarity.[4]

The variable p in Equation (5) refers to annual housing expenditure. Our sample not only includes home owners but also tenants. Following the urban literature (Gyourko and Tracy 1991; Bajari and Kahn 2005), we multiply the total housing value by 7.5% to convert it to the annual housing price for self-owned housings. For the rentals, the annual expenditure is calculated by multiplying the monthly rent by 12.

IPUMS follows the Public Use Micro Areas (PUMAs) to define the place of residence and assign community identifiers. PUMAs treat regions with population of 100,000+ people as communities within state lines. We use the community ID (i.e. PUMA) provided to calculate the community-level data, including *mCollege*, *mSpanish*, *mHindi*, *mChinese*, *mEnglish*, *mEnglishwhite*, and *mEnglishnonwhite*. We generate values for these new variables before dropping individual observations with missing values, to keep the information for each community accurate. Our final sample contains 13,699 observations with no missing values.

The sample is confined to households that moved in within the past four years.[5] The short period imposed guarantees that the housing demand information is updated and optimal decisions are made within recent years. Furthermore, in line with Bajari and Kahn (2005), we primarily include households wherein members drive private vehicles to work to isolate the impact of languages from that of public transportation on housing choice in this study, not to mention how complicated it is to measure the accessibility of public transit for each community and model the choice of commuting mode when choosing a housing unit.

Table 1 presents the summary statistics for household demographics and physical attributes of housing units by language group and city. The sample from Atlanta is composed of 86.42% English speakers, 9.67% Spanish speakers, 2.72% Hindi and related language speakers, and 1.19% Chinese speakers. The mother language distributions in Chicago and Dallas

[4]More specifically, we use data from Atlanta-Sandy Springs-Roswell (12,060 observations), Chicago-Naperville-Elgin (16,980 observations) and Dallas-Fort Worth-Arlington (19,100 observations), and we use Chicago, Atlanta, and Dallas for short in the paper.

[5]A robustness study has been conducted on the sample confined to households that moved in within the past nine years. The results are comparable and presented in Appendix (A1-A11).

Table 1. Summary statistics for household demographics and housing unit attributes.

Groups	Hexp1000	Hhincome1000	College	Famsize	Age	Married	Male	Rooms	Builtage	Single	Own
Atlanta(3,941)	20.137	111.150	0.576	2.455	37.997	0.457	0.501	6.153	29.175	0.554	0.488
Spanish(9.67%)	15.881	86.081	0.339	3.076	36.659	0.546	0.612	5.520	30.365	0.486	0.346
Hindi and related(2.72%)	25.214	129.503	0.841	3.065	39.822	0.813	0.785	5.888	19.299	0.505	0.570
Chinese(1.19%)	28.846	115.060	0.766	2.574	38.255	0.681	0.511	6.170	19.468	0.553	0.723
English(86.42%)	20.333	113.324	0.592	2.365	38.085	0.433	0.480	6.232	29.486	0.564	0.499
English-speaking white(53.16%)	23.696	135.961	0.679	2.370	38.094	0.510	0.532	6.603	29.716	0.642	0.616
English-speaking non-white(33.27%)	14.960	77.149	0.452	2.356	38.072	0.310	0.397	5.640	29.119	0.439	0.311
Chicago(4,129)	21.065	113.846	0.538	2.581	37.637	0.466	0.515	5.766	45.032	0.478	0.526
Spanish(15.28%)	14.854	78.449	0.235	3.330	37.101	0.539	0.531	5.403	51.355	0.471	0.463
Hindi and related(3.05%)	24.342	133.108	0.841	3.183	37.421	0.802	0.754	5.135	34.881	0.381	0.563
Chinese(1.26%)	23.132	101.800	0.692	2.846	38.481	0.750	0.712	5.615	35.885	0.365	0.615
English(80.41%)	22.089	120.031	0.581	2.412	37.733	0.436	0.500	5.861	44.358	0.485	0.535
English-speaking white(64.08%)	23.389	128.429	0.617	2.403	37.475	0.467	0.523	6.009	43.721	0.516	0.584
English-speaking non-white(16.32%)	16.985	87.061	0.442	2.445	38.748	0.312	0.411	5.282	46.860	0.364	0.340
Dallas(5,629)	20.073	109.458	0.517	2.471	37.368	0.469	0.528	5.451	28.539	0.504	0.403
Spanish(17.23%)	14.206	72.056	0.246	2.903	36.308	0.475	0.557	4.733	34.855	0.408	0.279
Hindi and related(2.58%)	25.316	133.618	0.903	3.076	37.566	0.793	0.828	5.814	18.110	0.503	0.497
Chinese(1.40%)	23.475	126.991	0.911	2.734	37.089	0.658	0.570	5.582	21.639	0.544	0.646
English(78.79%)	21.124	116.537	0.556	2.352	37.598	0.453	0.512	5.594	27.621	0.524	0.423
English-speaking white(61.18%)	22.455	125.090	0.585	2.380	37.768	0.492	0.534	5.829	27.779	0.583	0.481
English-speaking non-white(17.61%)	16.497	86.812	0.457	2.258	37.008	0.320	0.435	4.778	27.073	0.319	0.222

This table reports sample means for each city. Housing expenditure per year (Hexp1000) and annual household income (Hhincome1000) are measured in thousands of dollars. However, housing expenditure (p_j) used in the Hedonic function (see formula (5)) is measured in dollars.

are comparable, except that Chicago consists of slightly more English and Hindi and related language speakers and fewer Spanish and Chinese speakers. The sample sizes range from 3,941 to 5,629. Overall, the distributions of the languages spoken do not vary much across the three cities.

The English-speaker sample is further broken down into white and non-white. English-speaking non-white make up 33.27% in Atlanta, much higher than that in Chicago (16.32%) and Dallas (17.61%). Among English-speaking non-white groups, Blacks account for 85.29% in Atlanta, 65.03% in Chicago, and 70.45% in Dallas. On average, Chinese and Hindi and related language speakers are the top two groups with the highest income, education level, and marriage rate. Their housing units are among the newest. Spanish and English-speaking non-white groups are at the other end of the spectrum. English-speaking non-white in general own the fewest single-detached houses, and Chinese speakers prefer owning housings the most. Hindi and related languages and Spanish speakers are the most fertile groups.

Empirical motivation

We first run 10 separate parametric regressions (OLS or probit) for each city to obtain some stylized facts. Housing attributes and community level characteristics are the dependent variables. The explanatory variables include individual-level demographic characteristics and mother language information. The baseline results are shown in Tables 2–4.

The baseline group comprises white female English-speaker who are single and have no college education. In each table, columns 1–4 show the results for housing attributes. Compared with the baseline group, all the minority groups are less likely to live in a single detach house or own a housing. They live in places of residence with fewer rooms. Almost all the minority groups prefer newer buildings in all the cities, except Spanish speakers in Chicago and Dallas, along with English-speaking non-white groups in Chicago.

In all three cities, a higher income (measured by an increase of $1,000) is significant but it has a weak association with the community-level penetration rates of minority groups, whereas education has a much more significant correlation with the community characteristics. Communities differ in racial composition and human capital level. Compared with non-college graduates, college graduates tend to live in communities with 0.1%-0.3% more Hindi and related languages and Chinese speakers, or 0.38%-3.96% fewer Spanish and English-speaking non-white groups. This empirical evidence can be explained by the fact that Hindi and related languages and Chinese speakers are better educated, and they

Table 2. Descriptive OLS or probit regressions of housing choice – Atlanta.

Variables	(1) Rooms	(2) Builtage	(3) Single	(4) Own	(5) mSpanish	(6) mHindi	(7) mChinese	(8) mEnglishnonwhite	(9) mEnglishwhite	(10) mCollege
Hhincome1000	0.0062***	−0.0081**	0.0004***	0.0007***	−0.0000*	0.0000***	0.0000*	−0.0001***	0.0001***	0.0002***
	(0.0004)	(0.0032)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
College	0.2496***	−2.4417***	−0.0324**	0.1364***	−0.0066***	0.0025***	0.0011**	−0.0396***	0.0396***	0.0630***
	(0.0809)	(0.6716)	(0.0151)	(0.0151)	(0.0021)	(0.0004)	(0.0005)	(0.0058)	(0.0057)	(0.0043)
Famsize	0.6445***	0.0268	0.0892***	0.0216***	0.0002	−0.0002	−0.0002	−0.0003	0.0012	−0.0137***
	(0.0311)	(0.2583)	(0.0058)	(0.0058)	(0.0008)	(0.0002)	(0.0002)	(0.0022)	(0.0022)	(0.0016)
Age	0.0525***	−0.1005***	0.0079***	0.0074***	0.0000	0.0000	0.0000	−0.0005*	0.0005*	−0.0013***
	(0.0037)	(0.0308)	(0.0007)	(0.0007)	(0.0001)	(0.0000)	(0.0000)	(0.0003)	(0.0003)	(0.0002)
Married	0.7361***	−1.9586**	0.1956***	0.2288***	0.0073***	0.0006	−0.0001	−0.0326***	0.0236***	−0.0266***
	(0.0928)	(0.7706)	(0.0173)	(0.0173)	(0.0024)	(0.0005)	(0.0005)	(0.0067)	(0.0065)	(0.0049)
Male	0.0533	0.2538	0.0057	0.0274*	0.0003	0.0005	−0.0001	−0.0052	0.0032	0.0080**
	(0.0751)	(0.6233)	(0.0140)	(0.0140)	(0.0020)	(0.0004)	(0.0004)	(0.0054)	(0.0053)	(0.0039)
Spanish	−1.0970***	−0.6990	−0.2065***	−0.2022***	0.0466***	0.0056***	0.0040***	0.0216**	−0.1054***	−0.0159**
	(0.1322)	(1.0973)	(0.0246)	(0.0246)	(0.0034)	(0.0007)	(0.0008)	(0.0095)	(0.0093)	(0.0069)
Hindi and related	−1.4906***	−9.3888***	−0.2656***	−0.1672***	0.0168***	0.0094***	0.0073***	−0.0179	−0.0416***	0.0214*
	(0.2293)	(1.9032)	(0.0427)	(0.0427)	(0.0060)	(0.0013)	(0.0013)	(0.0165)	(0.0161)	(0.0120)
Chinese	−0.5886*	−9.8539***	−0.1301**	0.0666	0.0023	0.0087***	0.0095***	0.0033	−0.0454*	0.0453**
	(0.3377)	(2.8026)	(0.0629)	(0.0629)	(0.0088)	(0.0019)	(0.0019)	(0.0242)	(0.0238)	(0.0177)
English-speaking non-white	−0.3748***	−1.9843***	−0.1451***	−0.1816***	0.0104***	−0.0028***	−0.0010**	0.1256***	−0.1347***	−0.0473***
	(0.0852)	(0.7074)	(0.0159)	(0.0159)	(0.0022)	(0.0005)	(0.0005)	(0.0061)	(0.0060)	(0.0045)
Constant	1.6526***	37.1028***	−0.0064	−0.0371	0.0865***	0.0138***	0.0089***	0.3026***	0.5206***	0.4194***
	(0.1707)	(1.4165)	(0.0318)	(0.0318)	(0.0044)	(0.0009)	(0.0010)	(0.0123)	(0.0120)	(0.0090)
Observations	3,941	3,941	3,941	3,941	3,941	3,941	3,941	3,941	3,941	3,941
R-squared	0.3666	0.0261	0.2684	0.2767	0.0627	0.0762	0.0283	0.1755	0.1915	0.1980

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variables Single and Own are dummy variables; mSpanish, mHindi, mChinese, mEnglishnonwhite, mEnglishwhite and mCollege are community average variables; Rooms and Builtage are physical housing attributes. In each regression, the omitted category is female English-speaking white who are single and have not graduated from college. Household income is measured in thousands of dollars.

Table 3. Descriptive OLS or probit regressions of migrant housing choice -Chicago.

Variables	(1) Rooms	(2) Builtage	(3) Single	(4) Own	(5) mSpanish	(6) mHindi	(7) mChinese	(8) mEnglishnonwhite	(9) mEnglishwhite	(10) mCollege
Hhincome1000	0.0063***	−0.0248***	0.0006***	0.0008***	−0.0000*	−0.0000	0.0000	−0.0001***	0.0001**	0.0003***
	(0.0003)	(0.0038)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
College	0.1785**	−0.6951	−0.0653***	0.0670***	−0.0038	0.0034***	0.0024***	−0.0055	−0.0119**	0.0771***
	(0.0720)	(0.8082)	(0.0150)	(0.0155)	(0.0032)	(0.0008)	(0.0009)	(0.0042)	(0.0053)	(0.0045)
Famsize	0.5857***	−0.2832	0.1112***	0.0426***	−0.0027**	−0.0004	−0.0005	0.0013	0.0050**	−0.0139***
	(0.0269)	(0.3017)	(0.0056)	(0.0058)	(0.0012)	(0.0003)	(0.0003)	(0.0016)	(0.0020)	(0.0017)
Age	0.0353***	−0.0772**	0.0049***	0.0058***	−0.0002*	0.0000	−0.0001	0.0001	0.0000	−0.0012***
	(0.0033)	(0.0370)	(0.0007)	(0.0007)	(0.0001)	(0.0000)	(0.0000)	(0.0002)	(0.0002)	(0.0002)
Married	0.6031***	−1.4341	0.1677***	0.1894***	0.0003	0.0003	−0.0008	−0.0215***	0.0201***	−0.0081
	(0.0798)	(0.8949)	(0.0166)	(0.0172)	(0.0035)	(0.0009)	(0.0009)	(0.0047)	(0.0059)	(0.0050)
Male	0.0085	−0.7885	0.0464***	0.0594***	−0.0034	−0.0002	0.0006	−0.0080**	0.0133***	−0.0011
	(0.0657)	(0.7369)	(0.0137)	(0.0142)	(0.0029)	(0.0007)	(0.0008)	(0.0039)	(0.0048)	(0.0041)
Spanish	−0.7974***	6.4716***	−0.1565***	−0.1084***	0.0799***	0.0012	0.0062***	0.0258***	−0.1184***	0.0005
	(0.0976)	(1.0955)	(0.0204)	(0.0211)	(0.0043)	(0.0011)	(0.0012)	(0.0057)	(0.0072)	(0.0061)
Hindi and related	−1.6014***	−7.6901***	−0.2764***	−0.1494***	0.0017	0.0204***	0.0028	0.0058	−0.0684***	0.0245**
	(0.1900)	(2.1318)	(0.0396)	(0.0410)	(0.0084)	(0.0021)	(0.0023)	(0.0112)	(0.0140)	(0.0120)
Chinese	−0.7070**	−7.6869**	−0.2416***	−0.0428	0.0667***	0.0014	0.0526***	0.0168	−0.1385***	−0.0093
	(0.2888)	(3.2404)	(0.0603)	(0.0623)	(0.0127)	(0.0031)	(0.0034)	(0.0170)	(0.0213)	(0.0182)
English-speaking non-white	−0.4111***	1.7903*	−0.1209***	−0.1738***	0.0157***	−0.0025**	0.0043***	0.1548***	−0.1617***	−0.0266***
	(0.0906)	(1.0166)	(0.0189)	(0.0195)	(0.0040)	(0.0010)	(0.0011)	(0.0053)	(0.0067)	(0.0057)
Constant	2.0756***	51.9888***	−0.0684**	0.0027	0.1312***	0.0188***	0.0124***	0.1198***	0.6349***	0.3852***
	(0.1486)	(1.6677)	(0.0310)	(0.0321)	(0.0066)	(0.0016)	(0.0018)	(0.0087)	(0.0109)	(0.0093)
Observations	4,129	4,129	4,129	4,129	4,129	4,129	4,129	4,129	4,129	4,129
R-squared	0.3573	0.0445	0.2670	0.2164	0.0978	0.0347	0.0615	0.2025	0.1761	0.1831

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variables Single and Own are dummy variables; mSpanish, mHindi, mChinese, mEnglishnonwhite, mEnglishwhite and mCollege are community average variables; Rooms and Builtage are physical housing attributes. In each regression, the omitted category is female English-speaking white who are single and have not graduated from college. Household income is measured in thousands of dollars.

Table 4. Descriptive OLS or probit regressions of migrant housing choice – Dallas.

Variables	(1) Rooms	(2) Builtage	(3) Single	(4) Own	(5) mSpanish	(6) mHindi	(7) mChinese	(8) mEnglishnonwhite	(9) mEnglishwhite	(10) mCollege
Hhincome1000	0.0070***	−0.0114***	0.0006***	0.0009***	−0.0000	0.0000***	0.0000***	−0.0000***	0.0000**	0.0002***
	(0.0003)	(0.0027)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
College	0.2866***	−3.9502***	0.0334***	0.1199***	−0.0126***	0.0030***	0.0028***	−0.0039**	0.0080**	0.0485***
	(0.0610)	(0.5534)	(0.0119)	(0.0117)	(0.0025)	(0.0006)	(0.0004)	(0.0020)	(0.0038)	(0.0032)
Famsize	0.6172***	0.4445**	0.0975***	0.0539***	−0.0006	−0.0004	−0.0004**	−0.0006	0.0027*	−0.0142***
	(0.0240)	(0.2175)	(0.0047)	(0.0046)	(0.0010)	(0.0002)	(0.0002)	(0.0008)	(0.0015)	(0.0013)
Age	0.0467***	−0.0062	0.0076***	0.0073***	−0.0004***	0.0000	0.0000	−0.0001	0.0004**	−0.0007***
	(0.0028)	(0.0253)	(0.0005)	(0.0005)	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0002)	(0.0001)
Married	0.6821***	−1.3030**	0.2157***	0.2289***	−0.0131***	−0.0001	−0.0012***	−0.0026	0.0170***	−0.0103***
	(0.0708)	(0.6425)	(0.0138)	(0.0136)	(0.0029)	(0.0007)	(0.0005)	(0.0023)	(0.0044)	(0.0037)
Male	−0.0929	−0.6925	−0.0045	0.0142	0.0024	0.0005	0.0001	−0.0012	−0.0032	0.0034
	(0.0573)	(0.5201)	(0.0111)	(0.0110)	(0.0024)	(0.0006)	(0.0004)	(0.0018)	(0.0035)	(0.0030)
Spanish	−0.8673***	4.8883***	−0.1671***	−0.1293***	0.0591***	0.0004	−0.0008	0.0156***	−0.0814***	−0.0107**
	(0.0800)	(0.7265)	(0.0156)	(0.0154)	(0.0033)	(0.0008)	(0.0005)	(0.0026)	(0.0050)	(0.0042)
Hindi and related	−0.7652***	−8.0280***	−0.2255***	−0.1394***	−0.0032	0.0255***	0.0119***	0.0101*	−0.0931***	0.0431***
	(0.1793)	(1.6274)	(0.0348)	(0.0344)	(0.0074)	(0.0017)	(0.0012)	(0.0058)	(0.0111)	(0.0095)
Chinese	−0.6511***	−4.7486**	−0.1160**	0.0709	−0.0106	0.0246***	0.0187***	−0.0025	−0.0726***	0.0471***
	(0.2381)	(2.1616)	(0.0463)	(0.0457)	(0.0099)	(0.0023)	(0.0015)	(0.0077)	(0.0147)	(0.0126)
English-speaking non-white	−0.5266***	−1.8890***	−0.1821***	−0.1578***	0.0285***	0.0016**	−0.0003	0.0206***	−0.0606***	−0.0133***
	(0.0764)	(0.6933)	(0.0148)	(0.0147)	(0.0032)	(0.0007)	(0.0005)	(0.0025)	(0.0047)	(0.0040)
Constant	1.2619***	31.6992***	−0.1359***	−0.2219***	0.1743***	0.0138***	0.0101***	0.1600***	0.5740***	0.3793***
	(0.1231)	(1.1173)	(0.0239)	(0.0236)	(0.0051)	(0.0012)	(0.0008)	(0.0040)	(0.0076)	(0.0065)
Observations	5,629	5,629	5,629	5,629	5,629	5,629	5,629	5,629	5,629	5,629
R-squared	0.4197	0.0488	0.3452	0.3367	0.0942	0.0713	0.0697	0.0334	0.0900	0.1359

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variables Single and Own are dummy variables; mSpanish, mHindi, mChinese, mEnglishnonwhite, mEnglishwhite and mCollege are community average variables; Rooms and Builtage are physical housing attributes. In each regression, the omitted category is female English-speaking white who are single and have not graduated from college. Household income is measured in thousands of dollars.

may prefer to reside with others who speak the same language. More evidence is needed to support the latter hypothesis, and it will be discussed in the following section.

In Tables 2–4, columns 59 highlight the current residential segregation[6] or integration status. In Atlanta, all the minority groups live in communities with 4%-13% fewer English-speaking white compared with the baseline group. This may reflect a certain degree of segregation between English-speaking white and minorities. Based on this criterion, Chicago is the most segregated city, with the minorities living in communities with 6%-16% fewer English-speaking white. Dallas is in the middle, with the percentage ranging from 6% to 9%.

The baseline model provides some evidence on the determinants of residential segregation; however, it does not allow us to draw any causal conclusions because of the potential endogeneity issue. The endogeneity issue is a common empirical problem. Our approach originates from Bajari and Benkard (2005), which aims to address the endogeneity issue in the second stage of hedonic regression (Bartik 1987; Epple 1987). The advantage of the approach is that we can avoid estimating the second stage of the classic hedonic function (Rosen 1974) by assuming a linear utility function. Therefore, we do not need to estimate the whole market, and the household-level preference in our model can be locally identified. However, we still make the standard hedonic assumption that the unobserved product characteristics are independent of the observed product characteristics. We can recover only one single-dimension unobserved housing characteristic. Therefore, we must acknowledge that we cannot resolve all potential endogeneity problems. For example, some within-city amenities affect both housing decisions and preferences, which may still be excluded from our model and are beyond the capability of our model to capture. In the following section, we adopt the structural estimation model and estimate the willingness to pay of minority groups, allowing us to draw causal inferences.

IV. Structural model

Starting from the utility maximization problem, we first obtain the willingness to pay for each physical housing attribute and community human capital level, as measured by the percentage of college educated people. Next, we estimate the demand for the community-level percentage of different language speakers.

A key difference in interpreting the estimations for dichotomous and non-dichotomous variables needs to be addressed. For all the physical attributes that take dichotomous values, the estimates measure the willingness to pay for the jump from 0 to 1. Therefore, we can directly infer from the estimates how much a household would be willing to pay for an extra feature, maybe a room or a single detached unit. In the case of non-dichotomous variables, owing to the multi-valued or continuous nature embedded, we cannot interpret the coefficients in the same way. Instead, we estimate how much each household would be willing to pay for a fixed increase in any of the non-dichotomous variables, for example, an increase in the number of rooms from four to six or a 25% increase in the number of college educated residents or specific language speakers.

The ideal level of integration/segregation level varies across different language groups and cities. To assist interpreting the signs, an underlying assumption is imposed: all the language groups are presumed to have their own ideal level of integration/segregation relative to the baseline group, probably not the two extremes. The willingness to pay to live in a community with a higher percentage of people who speak a particular language, compared with the baseline group, not only depends on the current state of physical integration/segregation, but also is closely related to the household's preferred level of integration/segregation. A positive or negative

[6]OLS or Probit model are reduced-form regressions that can only infer correlations between attributes and residential choices of households.

sign does not directly indicate how severe or mild the current spatial segregation is; rather, it indicates how much segregation/integration (more or less) is desired by a household. A positive willingness to pay for a higher percentage of own language speakers means they desire more social and cultural support from others who speak their language. This reveals that the current percentage of own language speakers in the community is too low and a higher level of segregation is preferred. This is how we arrive at our conclusions about the present level of integration or segregation in the three cities. To sum up, no absolute measurement can be used to draw a conclusion on whether a community is over-integrated or over-segregated. We may only compare the current situation of minorities with their subjectively desired level, or, in other words, spatial segregation to preference segregation.

Willingness to pay for housing attributes and community characteristics

In this subsection, we summarize the results of willingness to pay for all sorts of physical housing attributes and community-level, non-language-related characteristics – the percentage of college educated individuals. The results are presented in Tables 5–9.

For all the non-dichotomous attributes or characteristics, we compute the marginal effects for a fixed increase, including increasing the number of rooms from four to six, building age from 10 to 35 years, and percentage of college educated households in the community from 10% to 35%. Conceptually, considering the number of rooms as an example, the marginal effect is defined as $WTP_{i,rooms} = \beta_{i,1}(log(6) - log(4))$ if the number of rooms increases from four to six. The other marginal effects can be defined in a similar way. For the other two dichotomous characteristics (single detached and ownership), computing the marginal effects is straightforward.

In all three cities, higher income, college educated, older, and married households with larger families are found to be unanimously willing to pay premiums for more rooms, newer building, and a single detached unit; to own a housing; and to live in a community with a higher level of human capital. For example, in Chicago, households are willing to pay $8.85 more per year with an extra $1,000 in income to increase the number of rooms from four to six, all else being equal. Married households are willing to pay $1,025.71 more than unmarried ones.

However, the willingness to pay for each housing attribute and the community-level percentage of college educated individuals varies across the different minority groups.

Compared with the baseline group, all the minority groups are unwilling to pay extra for two more rooms (Table 5). As shown in Table 6, Spanish and English-speaking non-white groups are more willing to pay extra for buildings that are 20 years or older in Chicago. Spanish, Hindi and related languages, and English-speaking non-white groups are willing to pay $355 – $734 more in Atlanta. It seems counter intuitive that minority groups, everything else being equal, are willing to pay more for older buildings. Considering newer residential properties are in general built in suburban areas and situated further from urban centres, positive willingness to pay for older buildings

Table 5. Estimates of the willingness to pay for more rooms.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	6.712***	8.853***	10.421***
	(0.368)	(0.377)	(0.406)
College	698.097***	471.600***	796.233***
	(76.520)	(80.986)	(82.960)
Famsize	433.251***	547.114***	538.770***
	(29.425)	(30.227)	(32.607)
Age	44.555***	36.346***	51.845***
	(3.514)	(3.707)	(3.795)
Married	1,139.769***	1,025.705***	1,451.160***
	(87.805)	(89.673)	(96.329)
Male	129.080*	228.136***	−39.292
	(71.021)	(73.834)	(77.978)
Spanish	−1,512.854***	−1,143.099***	−1,414.544***
	(125.026)	(109.773)	(108.920)
Hindi and related	−1,184.686***	−1,396.897***	−1,316.255***
	(216.847)	(213.615)	(243.969)
Chinese	205.881	−1,535.198***	−451.871
	(319.322)	(324.692)	(324.059)
English-speaking non-white	−1,408.531***	−1,334.402***	−1,333.777***
	(80.597)	(101.870)	(103.936)
Constant	528.779***	836.286***	267.654
	(161.387)	(167.103)	(167.497)
Observations	3,941	4,129	5,629
R-squared	0.448	0.430	0.431

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from four to six rooms in a housing unit ($(log(6) - log(4)) \cdot \beta_{i,1}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table 6. Estimates of the willingness to pay for an older building.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	−3.639***	−6.732***	−3.481***
	(0.201)	(0.285)	(0.195)
College	−459.558***	−747.110***	−274.983***
	(41.924)	(61.233)	(39.901)
Famsize	−146.390***	−186.873***	−108.984***
	(16.121)	(22.855)	(15.683)
Age	−19.463***	−15.942***	−11.280***
	(1.925)	(2.803)	(1.825)
Married	−502.281***	−628.591***	−195.021***
	(48.106)	(67.801)	(46.330)
Male	−58.221	−143.812**	25.877
	(38.911)	(55.826)	(37.504)
Spanish	596.763***	450.146***	−77.935
	(68.499)	(82.999)	(52.386)
Hindi and related	355.447***	217.394	−35.939
	(118.806)	(161.514)	(117.339)
Chinese	−310.768*	245.196	125.527
	(174.950)	(245.499)	(155.860)
English-speaking non-white	733.699***	1,094.786***	69.629
	(44.158)	(77.024)	(49.989)
Constant	−2,278.304***	−3,313.431***	−2,008.781***
	(88.421)	(126.346)	(80.560)
Observations	3,941	4,129	5,629
R-squared	0.386	0.374	0.154

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10 to 35 years in the age of a housing unit $((log(35) - log(10)) \cdot \beta_{i,2})$, holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table 7. Probit estimates of the demand for single detached houses.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	13.691***	15.52***	3.439***
	(1.414)	(1.384)	(0.547)
College	582.452**	−326.999	156.094
	(238.043)	(243.544)	(105.487)
Famsize	1334.228***	1720.024***	344.403***
	(123.284)	(131.829)	(44.129)
Age	124.457***	77.389***	24.909***
	(12.179)	(11.39)	(4.903)
Married	2771.502***	2472.003***	684.984***
	(287.336)	(273.839)	(122.359)
Male	189.675	869.004***	−42.434
	(219.013)	(219.531)	(98.649)
Spanish	−3828.97***	−2914.998***	−195.023
	(412.168)	(343.437)	(139.029)
Hindi and related	−3815.606***	−4241***	−389.844
	(681.854)	(676.693)	(314.98)
Chinese	−922.124	−5855***	−631.564
	(989.833)	(989.141)	(412.015)
English-speaking non-white	−3352.196***	−3500.998***	−374.446***
	(270.987)	(316.507)	(134.814)
Price of single detached unit	−1	−1	−1
Constant	−4367.751***	−4582.992***	−860.255***
	(633.625)	(677.449)	(237.858)
Pseudo-R2	0.438	0.43	0.0335
Observations	3941	4129	5629

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate probit estimate. The dependent variable is a dummy variable, "Single" equals 1 if the housing unit is single detached and 0 otherwise. *P*-values are reported in parentheses. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars, and the implicit price for single detached house is in dollars.

may suggest that households are truly willing to pay a premium for easier transit and a shorter commute. Perhaps they have a tendency to work in highly urbanized industries, a phenomenon known as 'job sprawl' (Bajari and Kahn 2005), and their job locations are systematically different from those of English-speaking white group. Nevertheless, we are unable to control for any commuting-related factors in the model since they are jointly determined with housing decisions. We provide some supporting evidence by conducting a *T*-test to demonstrate older residences are associated with shorter commutes. Based on the age of housing, households are categorized into two groups: building age undefined years and undefined years. We show that households who acquire older homes spend significantly less time commuting.

We extract the total amount of time in minutes it took a respondent to travel from home to work during the previous week from the data and perform the *T*-test. The average travel time to work is 28.691 minutes for those who acquired a home older than 20 years and 30.660 minutes for those who purchased a home less than 20 years. The difference is significant with a *T*-value of 5.161, as households who acquire newer homes commute for an additional 1.969 minutes.

In the case of single detached houses as depicted in Table 7, the willingness to pay is negative for all groups in Chicago, all groups except Chinese speakers in Atlanta and English-speaking non-white groups in Dallas. These findings lead to a tentative conclusion that minority groups are likely to locate in urban areas because housing units there tend to be smaller, older, and less likely to be single detached than those in suburban areas. Along similar lines, Bajari and Kahn (2005) explain urbanization of Blacks by showing their unwillingness to pay for two more rooms.

As depicted in Table 8, for housing ownership in general, English-speaking non-white and Hindi and related language speakers do not prefer to own a property in any of the three cities, similar to Spanish speakers in Atlanta and

Table 8. Probit estimates of the demand for ownership.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	8.852***	8.088***	9.438***
	(0.892)	(0.729)	(0.897)
College	1628.375***	1034.537***	1741.66***
	(151.457)	(121.367)	(165.342)
Famsize	188.024***	140.972***	112.177*
	(57.556)	(46.245)	(64.753)
Age	51.996***	21.084***	42.507***
	(7.089)	(5.596)	(7.779)
Married	1542.987***	860.545***	1436.109***
	(175.034)	(136.165)	(192.462)
Male	264.659*	335.948***	183.764
	(137.308)	(110.961)	(152.41)
Spanish	−1777.462***	39.453	−1102.452***
	(247.598)	(162.662)	(216.249)
Hindi and related	−1397.018***	−860.606***	−1179.127**
	(407.734)	(316.701)	(461.779)
Chinese	157.505	2279.415***	908.43
	(604.579)	(454.196)	(619.022)
English-speaking non-white	−2268.543***	−819.967***	−1659.048***
	(160.376)	(155.164)	(212.478)
Price of ownership	−1	−1	−1
Constant	234.871	−229.144	−360.452
	(349.181)	(268.702)	(403.462)
Pseudo-R2	0.384	0.219	0.209
Observations	3941	4129	5629

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate probit estimate. The dependent variable is a dummy variable, "Own" equals 1 if the migrants own a home and 0 if they rent a home. *P*-values are reported in parentheses. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars, and the implicit price for a single detached house is in dollars.

Table 9. Estimates of the willingness to pay for a higher percentage of college educated households in the community.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	36.517***	49.455***	46.473***
	(1.903)	(1.888)	(2.156)
College	4,582.958***	4,778.349***	3,905.428***
	(396.173)	(405.920)	(440.342)
Famsize	1,738.004***	1,505.781***	2,761.412***
	(152.345)	(151.506)	(173.075)
Age	196.942***	117.189***	285.054***
	(18.196)	(18.580)	(20.143)
Married	5,111.428***	4,436.881***	8,621.413***
	(454.598)	(449.460)	(511.301)
Male	898.027**	920.456**	130.243
	(367.701)	(370.073)	(413.897)
Spanish	−7,492.057***	−3,922.807***	−8,396.421***
	(647.308)	(550.210)	(578.132)
Hindi and related	−3,887.868***	−3,181.555***	−7,186.076***
	(1,122.698)	(1,070.691)	(1,294.954)
Chinese	3,392.400**	−4,390.082***	−3,163.218*
	(1,653.249)	(1,627.433)	(1,720.064)
English-speaking non-white	−7,417.844***	−6,593.625***	−7,027.718***
	(417.283)	(510.596)	(551.677)
Constant	5,604.903***	9,550.589***	−387.321
	(835.562)	(837.558)	(889.054)
Observations	3,941	4,129	5,629
R-squared	0.431	0.406	0.416

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are college graduates ($(log(35\%) - log(10\%)) \cdot \beta_{i,9}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Dallas. Chinese speakers prefer to own one, and they are willing to pay $2,279.41 more in Chicago. The estimates are positive but insignificant in the other two cities.

The results for willingness to pay obtained from the structural model vary from the benchmark OLS/probit estimations. This further confirms that the endogeneity issue might be more problematic in the reduced form regressions.

Finally, we estimate the minorities' willingness to pay for an increase in the percentage of college educated individuals in the community from 10% to 35%, holding all else equal. Table 9 depicts that all the minority groups have negative willingness to pay in all three cities except for Chinese speakers in Atlanta, in line with the finding for Blacks documented in Bajari and Kahn (2005).

All the regression results have been interpreted under the assumption that all else is equal and we only allow one variable to change at a time. We recognize that demographic and taste differences jointly determine the demand for the level of human capital in the community, and all else is not being equal.

According to Table 1, Hindi and related languages and Chinese speakers are the minority groups with the highest level of education among all the minority groups, while Spanish and English-speaking non-white groups have the lowest average education level. The negative willingness to pay for higher degree of college education level in the community suggests that all minority groups engage in sorting behaviours. Speakers of Hindi and related languages, and Chinese already reside in communities with a high ratio of college graduates, and they would not wish to pay more for further increments. In contrast, Spanish- and English-speaking non-white groups choose to live in urban districts with smaller proportion of college graduates relative to the suburbs during the urbanization process, as found by Bajari and Kahn (2005).

To summarize our findings on all the minority groups except Chinese speakers, the preferences for fewer rooms, older buildings, non-single detached units, non-ownership, and fewer college educated individuals in the community provide strong evidence of their enthusiasm for urbanization. For the Chinese speakers, the findings are mixed. Considering that the Chinese speakers are composed of Mandarin and Cantonese speakers, the huge heterogeneity between the two language subgroups complicates the interpretation of the estimates.

Demand for integration and segregation

This section focuses on integration and segregation, specifically the extent to which minority groups are willing to become more integrated or separated, i.e. the gap between spatial and preference segregation.

Demand for integration

We measure the demand for a higher level of integration, that is, we estimate household i's willingness to pay for increasing the percentage of the English-speaking white group in the community from 10% to 35%, denoted by $WTP_{i,Englishwhite}$. We apply the random coefficient estimation strategy by assuming that all other groups decline proportionally.

$$\begin{aligned} WTP_{i,Englishwhite} = {} & \beta_{i,8} \times (\log(35\%) - \log(10\%)) \\ & + \beta_{i,5} \times (\log(65\% \times R_{i,Spanish}) - \log(90\% \times R_{i,Spanish})) \\ & + \beta_{i,6} \times (\log(65\% \times R_{i,Hindi}) - \log(90\% \times R_{i,Hindi})) \\ & + \beta_{i,7} \times (\log(65\% \times R_{i,Chinese}) - \log(90\% \times R_{i,Chinese}))cr \end{aligned}, \quad (11)$$

where $R_{i,Spanish}$, $R_{i,Hindi}$, $R_{i,Chinese}$ and $R_{i,Englishnonwhite}$ denote the initial proportions of each minority group in each IPUM community. Without loss of generality, we assume that by increasing the proportion of English-speaking white, the percentage of each minority group in the community is reduced by an equal amount. Similarly, by estimating the willingness to pay for an increase of 25% of any one of the minority groups $WTP_{i,Spanish}$, $WTP_{i,Hindi}$, $WTP_{i,Chinese}$ and $WTP_{i,Englishnonwhite}$, we evaluate the demand for residential segregation.

Next, we regress the willingness to pay for an increase in the proportion of English-speaking white (from 10% to 35%) in the community on household demographics. Table 10 presents the results for the three cities.

In Chicago and Dallas, higher income, college education, larger family size, older age, married status, and being male all show negative impacts on the willingness to pay for living in a community that has 35% of the English-speaking white versus a community that has 10% of the English-speaking white (holding everything else unchanged). In Atlanta, the demographic effects are all positive.

The minority groups exhibit different preferences for living with more English-speaking white. In Chicago, all the minority households are willing to pay more to live in a whiter English-speaking community, reflecting a unanimous taste for a higher level of integration. The level of spatial segregation exceeds that of preference segregation, and they desire to be more integrated.

In Atlanta, Spanish, Hindi and related languages and English-speaking non-white groups are willing

Table 10. Estimates of the willingness to pay for more English-speaking white.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	25.712***	−29.557***	−22.577***
	(1.286)	(1.383)	(1.291)
College	2,593.498***	−1,473.808***	−2,595.174***
	(267.736)	(297.482)	(263.704)
Famsize	962.009***	−1,651.297***	−1,030.044***
	(102.956)	(111.033)	(103.648)
Age	84.881***	−94.643***	−124.810***
	(12.297)	(13.617)	(12.063)
Married	2,362.402***	−3,153.536***	−3,823.226***
	(307.220)	(329.391)	(306.199)
Male	653.693***	−1,000.501***	−259.187
	(248.494)	(271.211)	(247.867)
Spanish	−4,307.660***	4,917.733***	2,804.992***
	(437.455)	(403.227)	(346.221)
Hindi and related	−2,305.403***	3,507.844***	−1,366.485*
	(758.726)	(784.666)	(775.498)
Chinese	2,365.332**	4,964.249***	−1,889.101*
	(1,117.276)	(1,192.679)	(1,030.081)
English-speaking non-white	−3,687.384***	5,771.022***	1,523.528***
	(282.002)	(374.195)	(330.378)
Constant	−2,594.596***	−1,373.644**	5,416.685***
	(564.678)	(613.812)	(532.421)
Observations	3,941	4,129	5,629
R-squared	0.351	0.379	0.285

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are English-speaking white (that is, $WTP_{i,Englishwhite}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Both housing expenditure per year and annual household income are measured in thousands of dollars.

to pay less compared to the English-speaking white to live in a community with 25% more English-speaking white, showing no interest in more integration. Spanish and English-speaking non-white groups are willing to pay $4,308 and $3,687 less, respectively. The magnitudes are large enough to offset the positive income and education effects, which have minimal impacts on these two minority groups per se, as they are the least paid and the least educated groups. Such large negative willingness to pay indicates that the level of preference segregation is a lot lower than their current spatial segregation for them in Atlanta. In contrast, Chinese speakers are willing to pay $2,365 more. Considering that the Chinese are one of the best educated groups, the positive language effect strengthens the education effect, implying a very strong preference for a higher level of integration.

In Dallas, only Spanish and English-speaking non-white groups show interests in greater integration. The same observation applies to Hindi and related languages and Chinese speakers. The education effect intensifies the language effect, and both groups show strong reluctance towards increased integration. However, Spanish and English-speaking non-white groups are willing to pay $2,805 and $1,524 more, respectively, which is attenuated partially by the negative education effect. Overall, Dallas shows a moderate level of preference for higher level of integration.

Across the three cities, Chicago is the only city where all the minority groups show a strong preference for more integration, implying that it has the lowest level of integration. Alternatively, it is the most segregated city among the three, consistent with the 1990 dissimilarity index (Cutler, Glaeser, and Vigdor 1999).

So far, we have shown that minority groups are faced with different levels of segregation/integration in the three cities. Next, we examine how the willingness to integrate into or segregate from the English-speaking white is related to minority groups' preferences for living in communities with their own language speakers and other minority language groups. We further attempt to trace the origins of integration or segregation and provide some evidence to support our hypothesis that spoken language plays a key role in determining residential location and shaping the spatial segregation landscape.

Choosing to segregate?

As stated in the introduction, members of a minority group may prefer speaking their mother language at home or in the community for the sake of daily convenience and sense of belonging. The benefits lie not only in the easiness of expression, but also in the accuracy and depth of the conversation. Speaking the same language implies sharing similar culture, religion, history, social norms, and food habits and, thus, common values. This provides incentives for minorities to choose to live in a community with a high penetration of people who speak the same language.

We conduct four separate tests, each by increasing the percentage of one minority language speaker from 10% to 35% in the community. We are interested in evaluating the willingness to pay for own language speakers, and the willingness to pay for other minority language speakers. By comparing these two sets of preferences with the preference for a greater proportion of white English-speakers, we can draw inferences regarding segregation.

Table 11. Estimates of the willingness to pay for More Spanish speakers.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	1.270***	7.544***	20.477***
	(0.228)	(0.350)	(0.823)
College	−89.066*	499.319***	1,723.762***
	(47.443)	(75.343)	(168.001)
Famsize	230.455***	405.305***	1,023.094***
	(18.244)	(28.121)	(66.032)
Age	28.769***	28.590***	106.029***
	(2.179)	(3.449)	(7.685)
Married	595.885***	834.626***	3,221.335***
	(54.439)	(83.424)	(195.074)
Male	23.631	237.365***	30.716
	(44.033)	(68.689)	(157.912)
Spanish	−266.417***	−672.313***	−2,630.915***
	(77.517)	(102.124)	(220.572)
Hindi and related	−152.958	−695.237***	−2,040.300***
	(134.446)	(198.731)	(494.056)
Chinese	555.086***	−478.425	−1,010.213
	(197.981)	(302.067)	(656.246)
English-speaking non-white	−236.189***	−626.783***	−2,295.797***
	(49.971)	(94.771)	(210.478)
Constant	290.910***	277.628*	−2,058.963***
	(100.061)	(155.459)	(339.196)
Observations	3,941	4,129	5,629
R-squared	0.237	0.355	0.428

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who speak Spanish at home (that is, $WTP_{i,Spanish}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table 12. Estimates of the willingness to pay for more Hindi and related language speakers.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	−20.488***	4.789***	−2.521***
	(1.065)	(0.326)	(0.611)
College	−2,281.688***	58.226	310.928**
	(221.730)	(70.019)	(124.733)
Famsize	−932.136***	403.962***	−84.038*
	(85.265)	(26.134)	(49.026)
Age	−90.021***	22.069***	2.848
	(10.184)	(3.205)	(5.706)
Married	−2,529.376***	792.029***	56.403
	(254.430)	(77.529)	(144.833)
Male	−520.235**	193.587***	62.486
	(205.795)	(63.835)	(117.242)
Spanish	4,169.554***	−911.168***	−389.094**
	(362.286)	(94.908)	(163.764)
Hindi and related	2,151.596***	−847.957***	1,946.585***
	(628.353)	(184.688)	(366.814)
Chinese	−1,908.933**	−1,999.876***	1,403.465***
	(925.292)	(280.722)	(487.233)
English-speaking non-white	3,771.010***	−1,479.718***	632.888***
	(233.545)	(88.075)	(156.270)
Constant	1,980.731***	694.587***	−439.001*
	(467.648)	(144.474)	(251.837)
Observations	3,941	4,129	5,629
R-squared	0.398	0.322	0.017

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who speak Hindi and related languages at home (that is, $WTP_{i,Hindi}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table 13. Estimates of the willingness to pay for more Chinese speakers.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	−9.954***	8.954***	15.963***
	(0.575)	(0.512)	(0.792)
College	−1,338.565***	335.616***	1,656.559***
	(119.773)	(110.065)	(161.728)
Famsize	−214.328***	579.985***	526.658***
	(46.058)	(41.081)	(63.567)
Age	−17.464***	32.637***	64.919***
	(5.501)	(5.038)	(7.398)
Married	−581.271***	883.628***	2,064.334***
	(137.436)	(121.871)	(187.789)
Male	−260.263**	365.189***	174.043
	(111.165)	(100.345)	(152.015)
Spanish	1,675.089***	−1,782.401***	−1,630.320***
	(195.698)	(149.190)	(212.335)
Hindi and related	228.288	−1,374.204***	1,341.934***
	(339.420)	(290.319)	(475.607)
Chinese	−1,409.159***	−1,119.403**	1,986.013***
	(499.819)	(441.279)	(631.741)
English-speaking non-white	2,084.078***	−1,768.701***	−885.205***
	(126.155)	(138.448)	(202.618)
Constant	811.577***	−538.745**	−2,764.077***
	(252.612)	(227.104)	(326.530)
Observations	3,941	4,129	5,629
R-squared	0.299	0.303	0.283

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who speak Chinese at home (that is, $WTP_{i,Chinese}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

We run separate estimations of Equation (11) for Spanish, Hindi and related languages, Chinese speakers, and non-white English-speakers to obtain the willingness to pay for 25% more of each minority language speaker in the community. The results are shown in Tables 11–14; Two patterns emerge as we carefully examine the willingness to pay across cities and groups.

First, for each minority language group with significant estimations, the willingness to pay for own language speakers and white English-speaker are largely negatively correlated. For example, in Atlanta and Chicago, Chinese speakers are willing to pay $2,365 and $4,964 more, respectively, to live with 25% more English-speaking white; however, they would pay $1,409 and $1,119 less to live with 25% more Chinese speakers. In Dallas, Chinese speakers would pay $1,889 less to live with more English-speaking white, but they would pay $1,986 more to live with more Chinese speakers. In Atlanta and Dallas, Hindi and related language speakers would pay $2,305 and $1,366 less, respectively, to have more English-speaking white, but they would pay $2,152 and $1,947 more, respectively, to have more own language speakers. For Hindi and related language speakers in Chicago, the willingness to pay for more English-speaking white and own language speakers is $3,508 and -$848, respectively. This implies that the demands for English-speaking white and own language speakers become substitutes for each other. The only exceptions are English-speaking non-white groups in Atlanta and Dallas. The signs for the willingness to pay are consistent. The mixed results may arise from the complicated composition of English-speaking non-white groups. Although 70%-80% of the English-speaking non-white groups are African American, there are significant percentage of other minorities mixed in this group.

Second, any two minority groups share the same reciprocal attitudes towards one another. In Atlanta, for instance, Spanish speakers have no desire in having more English-speaking non-white groups in the community, as

Table 14. Estimates of the willingness to pay for more English-speaking non-white groups.

Variables	Atlanta	Chicago	Dallas
Hhincome1000	1.283***	10.100***	−9.249***
	(0.407)	(0.483)	(0.447)
College	837.936***	682.021***	−852.752***
	(84.741)	(103.910)	(91.305)
Famsize	−113.838***	364.068***	−353.162***
	(32.586)	(38.783)	(35.887)
Age	−12.763***	17.699***	−38.554***
	(3.892)	(4.756)	(4.177)
Married	−52.710	855.232***	−1,199.428***
	(97.238)	(115.055)	(106.018)
Male	49.653	264.232***	14.798
	(78.651)	(94.733)	(85.822)
Spanish	−874.176***	−1,804.990***	1,544.197***
	(138.459)	(140.846)	(119.876)
Hindi and related	234.966	−808.259***	310.101
	(240.144)	(274.082)	(268.509)
Chinese	158.629	−1,674.610***	−227.892
	(353.628)	(416.600)	(356.656)
English-speaking non-white	−1,500.361***	−2,266.186***	887.382***
	(89.256)	(130.705)	(114.390)
Constant	−348.808*	941.064***	−581.929***
	(178.726)	(214.403)	(184.346)
Observations	3,941	4,129	5,629
R-squared	0.144	0.336	0.308

Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are non-white and speak English at home (that is, $WTP_{i,Englishnonwhite}$), holding all other housing attributes constant. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

evidenced by their -$874 willingness to pay. And, the English-speaking non-white groups do not prefer to live in communities with more Spanish speakers either, with the willingness to pay being $236 less. In Dallas, Chinese speakers are interested in living in communities with more Hindi and related language speakers and are willing to pay $1,403 more; Hindi and related language speakers have a similar attitude towards Chinese speakers and are willing to pay $1,342 more to live in a community with more Chinese speakers.

A corollary inference can be formed from the discovered patterns. If a minority group wishes to become more integrated, they desire to have fewer speakers of its own language, but may not prefer to reduce the proportion of other minority groups. To be segregated from the mainstream, a minority group would want to dwell in a neighbourhood with more people who speak their own language.

V. Further discussion

We have depended on a rough classification of spoken languages up until this point. We further subdivide a broad category of language group into finer divisions and conduct several tests. We employ Chinese-speaking individuals to elaborate our reasoning. Same reasoning can be applied to the Hindi and related languages speakers.

The willingness of Chinese speakers to pay more to reside in a community with 25% more speakers of their own language reflects the average preference of Cantonese and Mandarin speakers. If our hypothesis on language as a determinant of integration or segregation is correct, we anticipate that doing separate tests for Cantonese and Mandarin speakers will amplify the effect. If Chinese speakers demonstrate a preference for more Chinese, we believe that each subgroup will be more eager to have speakers of their own subdivided language present, thus embracing the greater convenience and sense of belonging.

Table 15. WTP for Hindi and related language speakers versus WTP for Hindi speakers.

Variables	Chicago		Dallas	
	WTP for Hindi and Related	WTP for Hindi-only	WTP for Hindi and Related	WTP for Hindi-only
Hindi and Related	−847.957***		1,946.585***	
	(184.688)		(366.814)	
Hindi-only		−1,353.624***		5,163.574***
		(343.203)		(886.155)
Related		−440.607		3,560.636***
		(273.532)		(800.966)

This table shows the willingness to pay (WTP) for a higher percentage of Hindi and related language speakers and a higher percentage of Hindi-only speakers per se, respectively. Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. The omitted category is female English-speaking white who are single and have not graduated from college. Annual household income is measured in thousands of dollars, and the implicit price for a single detached house is in dollars.

Similarly, if Chinese speakers resist having more Chinese in their communities, we anticipate that the negative feelings will intensify.

Chinese and Hindi and related languages provide excellent test cases for our idea. However, due to the problem of missing data, there are insufficient samples to conduct empirical tests for Chinese speakers, and the only alternative is to test Hindi and related language speakers. We divide them into two groups based on language classification by separating Hindi from the rest of the category and conduct separate estimates for each group.

Table 15 depicts the willingness to pay for a 25% increase in Hindi-only speakers in the community, relative to Table 12's results. We find that the Hindi group's willingness to pay for more of Hindi-only group is -$1,354 in Chicago and $5,164 in Dallas, both of which are significantly greater than the average willingness to pay for an increase in speakers of Hindi and related language. This supports our prediction based on the hypothesis. The finer the classification of a language, the greater the language effects.

VI. Conclusion

This paper analysed the housing choices of minority groups in order to examine their assimilation into mainstream white English-speaking community. In general, the findings indicate that minorities do not prioritize larger rooms, single-family detached homes, or home ownership. They were willing to pay extra for older structures in Chicago and Dallas. With the exception of Chinese speakers, the estimations on housing attributes largely explain the urbanization and segregation of minority groups.

We found that Chicago is the most segregated of the three sample cities. Unlike the classic dissimilarity index, our assessment of the amount of segregation is an indirect technique that discloses each minority group's subjective feelings towards segregation. We discovered evidence that individuals unwilling to pay to live with more English-speaking whites choose segregation to enjoy the comfort of oral communication and to share a wide range of common values. In addition, we found evidence that the demands for more English-speaking whites and own-language speakers are direct substitutes for achieving a desired level of integration or segregation.

We further demonstrated the significance of language by subdividing the speakers of Hindi and related languages into finer language subgroups. We found that a finer categorization of spoken language results in greater magnitudes of willingness to pay in Chicago and Dallas, offering additional evidence supporting the role of spoken language as a determinant of residential choice, hence shaping spatial segregation.

Notation and description of variables

Notation	Description
Rooms	number of rooms
Builtage	age of the residence (years)
Single	single-detached housing unit (1=yes, 0=no)
Own	ownership of the the housing unit (1=own, 0=rent)
mSpanish	percentage of Spanish speakers
mHindi	percentage of Hind and related language speakers
mChinese	percentage of Chinese speakers
mEnglishwhite	percentage of English-speaking white respondents
mEnglishnonwhite	percentage of English-speaking non-white respondents
mCollege	percentage of college-educated respondents
ξ	the unobserved attribute of the housing unit
c	consumption of a composite commodity
Hhincome1000	household's total annual income (in 1,000 dollars)
College	college-graduated (1=yes, 0=no)
Famsize	family size
Age	the age of the household head
Married	marital status (1=married, 0=not married)
Male	gender (1=male, 0=female)
Spanish	speak Spanish at home (1=yes, 0=no)
Hindi and related	speak Hindi and related language at home (1=yes, 0=no)
Chinese	speak Chinese at home (1=yes, 0=no)
English-speaking non-white	English-speaking non-white (1=yes, 0=no)
English-speaking white	English-speaking white (1=yes, 0=no)

mSpanish, mHindi, mChinese, mEnglishwhite, mEnglishnonwhite, and mCollege are variables at the community level.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study is supported by the National Natural Science Foundation of China [Project No.'s: 71828301 and 71673166].

ORCID

Tiantian Dai http://orcid.org/0000-0001-6019-9152
Buyuan Yang http://orcid.org/0000-0002-4083-0194

References

Allen, R., S. Burgess, R. Davidson, and F. Windmeijer. 2015. "More Reliable Inference for the Dissimilarity Index of Segregation." *The Econometrics Journal* 18 (1): 40–66. doi:10.1111/ectj.12039.

Alonso-Villar, O., C. Gradin, and C. Del Río. 2013. "Occupational Segregation of Hispanics in Us Metropolitan Areas." *Applied Economics* 45: 4298–4307. doi:10.1080/00036846.2013.778951.

Bajari, P., and C. L. Benkard. 2005. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." *The Journal of Political Economy* 113: 1239–1276. doi:10.1086/498586.

Bajari, P., and M. E. Kahn. 2005. "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities." *Journal of Business & Economic Statistics* 23: 20–33. doi:10.1198/073500104000000334.

Bartik, T. J. 1987. "The Estimation of Demand Parameters in Hedonic Price Models." *The Journal of Political Economy* 95: 81–88. doi:10.1086/261442.

Bayer, P., R. McMillan, and K. Rueben, 2004a. "*An Equilibrium Model of Sorting in an Urban Housing Market*." Technical Report. National Bureau of Economic Research.

Bayer, P., R. McMillan, and K. Rueben. 2004b. "What Drives Racial Segregation? New Evidence Using Census Microdata." *Journal of Urban Economics* 56: 514–535. doi:10.1016/j.jue.2004.06.002.

Bayer, P., F. Ferreira, and R. McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *The Journal of Political Economy* 115: 588–638. doi:10.1086/522381.

Bayer, P., and R. McMillan. 2012. "Tiebout Sorting and Neighborhood Stratification." *Journal of Public Economics* 96: 1129–1143. doi:10.1016/j.jpubeco.2012.02.006.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins. 2016. "A Dynamic Model of Demand for Houses and Neighborhoods." *Econometrica* 84: 893–942. doi:10.3982/ECTA10170.

Bolt, G., D. Phillips, and R. V. Kempen. 2010. "Housing Policy, (de)segregation and Social Mixing: An International Perspective." *Housing Studies* 25: 129–135. doi:10.1080/02673030903564838.

Borjas, G. J. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *The American Economic Review* 85: 365–390.

Conejeros, R., and M. Vargas. 2012. "Segregation, Exclusion and Compensating Transfers Under a Dynamic Setting." *Applied Economics* 44: 2203–2215. doi:10.1080/00036846.2011.562171.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *The Journal of Political Economy* 107: 455–506. doi:10.1086/250069.

De la Roca, J., I. G. Ellen, and K. M. O'-Regan. 2014. "Race and Neighborhoods in the 21st Century: What Does Segregation Mean Today?" *Regional Science and Urban Economics* 47: 138–151. doi:10.1016/j.regsciurbeco.2013.09.006.

Epple, D. 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *The Journal of Political Economy* 95: 59–80. doi:10.1086/261441.

Epple, D., M. Peress, and H. Sieg. 2010. "Identification and Semiparametric Estimation of Equi- Librium Models of Local Jurisdictions." *American Economic Journal: Microeconomics* 2: 195–220.

Fan, J., and I. Gijbels. 2018. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*. Routledge.

Gilroy, P. 2013. *Between Camps: Nations, Cultures and the Allure of Race*. Routledge.

Gradín, C. 2020. "Segregation of Women into Low-Paying Occupations in the United States." *Applied Economics* 52: 1905–1920. doi:10.1080/00036846.2019.1682113.

Gyourko, J., and J. S. Tracy. 1991. "The Structure of Local Public Finance and the Quality of Life." *The Journal of Political Economy* 99: 774–806. doi:10.1086/261778.

Kain, J. F. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization." *The Quarterly Journal of Economics* 82: 175–197. doi:10.2307/1885893.

Koster, H. R. A., J. Van Ommeren, and P. Rietveld. 2016. "Historic Amenities, Income and Sorting of Households." *Journal of Economic Geography* 16: 203–236. doi:10.1093/jeg/lbu047.

Kuminoff, N. V., V. K. Smith, and C. Timmins. 2013. "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets." *Journal of Economic Literature* 51: 1007–1062. doi:10.1257/jel.51.4.1007.

Massey, D. S., and N. A. Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." *American Sociological Review* 52: 802. doi:10.2307/2095836.

Massey, D. S., and N. A. Denton. 1989. "Hypersegregation in U.S. Metropolitan Areas: Black and Hispanic Segregation Along Five Dimensions." *Demography* 26: 373–391. doi:10.2307/2061599.

Moulin, L., D. Flacher, and H. Harari-Kermadec. 2016. "Tuition Fees and Social Segregation: Lessons from a Natural Experiment at the University of Paris 9-Dauphine." *Applied Economics* 48: 3861–3876. doi:10.1080/00036846.2016.1148253.

Peach, C. 1996. "Good Segregation, Bad Segregation." *Planning Perspectives* 11: 379–398. doi:10.1080/026654396364817.

Phillips, D. 2006. "Parallel Lives? Challenging Discourses of British Muslim Self-Segregation." *Environment and Planning D, Society & Space* 24: 25–40. doi:10.1068/d60j.

Phillips, D., C. Davis, and P. Ratcliffe. 2007. "British Asian Narratives of Urban Space." *Transac- Tions of the Institute of British Geographers* 32: 217–234. doi:10.1111/j.1475-5661.2007.00247.x.

Phillips, D. 2007. "Ethnic and Racial Segregation: A Critical Perspective." *Geography Compass* 1: 1138–1159. doi:10.1111/j.1749-8198.2007.00051.x.

Reardon, S. F., and D. O'-Sullivan. 2004. "Measures of Spatial Segregation." *Sociological Methodology* 34: 121–162. doi:10.1111/j.0081-1750.2004.00150.x.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Com- Petition." *The Journal of Political Economy* 82: 34–55. doi:10.1086/260169.

Sieg, H., V. K. Smith, H. S. Banzhaf, and R. P. Walsh. 2004. "Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods." *International Economic Review* 45: 1047–1077. doi:10.1111/j.0020-6598.2004.00297.x.

Wilson, W. J. 2012. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

Wong, M. 2013. "Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore." *The Review of Economic Studies* 80: 1178–1214. doi:10.1093/restud/rdt002.

Yalonetzky, G. 2012. "A Dissimilarity Index of Multidimensional Inequality of Opportunity." *Journal of Economic Inequality* 10: 343–373. doi:10.1007/s10888-010-9162-z.

Zhang, J., and L. Zheng. 2015. "Are People Willing to Pay for Less Segregation? Evidence from U.S. Internal Migration." *Regional Science and Urban Economics* 53: 97–112. doi:10.1016/j.regsciurbeco.2015.05.002.

Appendix

Table A1. Summary statistics for household demographics and housing unit attributes.

GROUPS	Hexp1000	Hhincome1000	College	Famsize	Age	Married	Male	Rooms	Builtage	Single	Own
Atlanta (5053)	21.080	116.639	0.575	2.569	39.317	0.489	0.507	6.485	29.240	0.608	0.547
Spanish (9.40%)	15.832	85.187	0.328	3.229	37.745	0.554	0.613	5.699	30.402	0.533	0.389
Hindi and related (2.65%)	25.916	133.581	0.821	3.343	40.485	0.836	0.784	6.284	19.187	0.575	0.619
Chinese (1.35%)	31.200	126.909	0.706	2.853	40.294	0.735	0.500	6.559	20.397	0.662	0.779
English (86.60%)	21.344	119.375	0.592	2.469	39.436	0.468	0.487	6.575	29.560	0.616	0.558
English-speaking white (53.93%)	25.114	142.681	0.673	2.498	39.488	0.551	0.545	6.992	29.897	0.694	0.670
English-speaking non-white (32.67%)	15.122	80.907	0.457	2.422	39.351	0.331	0.392	5.887	29.003	0.488	0.372
Chicago (5548)	21.357	119.553	0.532	2.705	38.980	0.506	0.523	6.000	45.056	0.526	0.587
Spanish (15.70%)	14.692	77.980	0.222	3.440	38.277	0.563	0.546	5.540	51.866	0.505	0.506
Hindi and related (2.99%)	24.723	126.841	0.801	3.428	38.717	0.831	0.753	5.404	35.229	0.440	0.627
Chinese (1.53%)	25.108	119.299	0.753	2.906	40.824	0.788	0.682	5.682	38.847	0.447	0.753
English (79.78%)	22.471	127.466	0.579	2.529	39.093	0.477	0.507	6.119	44.204	0.535	0.598
English-speaking white (64.46%)	23.757	136.036	0.612	2.532	38.908	0.510	0.529	6.284	43.650	0.569	0.650
English-speaking non-white (15.32%)	17.060	91.411	0.441	2.518	39.871	0.341	0.414	5.426	46.534	0.389	0.378
Dallas (7153)	20.787	116.150	0.519	2.585	38.802	0.505	0.534	5.722	29.169	0.565	0.473
Spanish (17.45%)	14.473	76.080	0.242	3.030	37.641	0.509	0.572	4.923	35.534	0.475	0.351
Hindi and related (2.61%)	26.130	149.152	0.904	3.214	38.872	0.824	0.824	6.267	18.080	0.578	0.588
Chinese (1.43%)	23.538	131.844	0.912	2.765	38.804	0.657	0.578	5.706	22.730	0.569	0.676
English (78.51%)	21.962	123.671	0.561	2.462	39.058	0.491	0.515	5.882	28.241	0.585	0.492
English-speaking white (62.35%)	23.334	132.677	0.586	2.500	39.387	0.530	0.537	6.120	28.459	0.643	0.552
English-speaking non-white (16.16%)	16.671	88.922	0.463	2.315	37.788	0.343	0.431	4.965	27.398	0.359	0.261

This table reports sample means for each city. Housing expenditure per year (Hexp1000) and annual household income (Hhincome1000) are measured in thousands of dollars.

Table A2. Estimates of the willingness to pay for more rooms.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	7.474***	7.827***	10.568***
	(0.363)	(0.311)	(0.367)
College	1,062.368***	660.684***	941.192***
	(80.757)	(72.434)	(79.165)
Famsize	477.429***	540.049***	553.848***
	(29.942)	(26.487)	(30.557)
Age	58.634***	45.038***	62.187***
	(3.695)	(3.318)	(3.598)
Married	1,252.717***	1,107.780***	1,530.924***
	(92.623)	(80.652)	(92.056)
Male	182.016**	224.353***	−67.820
	(75.495)	(66.075)	(74.244)
Spanish	−1,763.163***	−1,182.933***	−1,583.931***
	(133.940)	(97.350)	(103.154)
Hindi and related	−1,391.749***	−1,447.435***	−1,485.968***
	(231.898)	(192.754)	(230.109)
Chinese	358.006	−1,442.226***	−767.709**
	(318.497)	(263.774)	(305.708)
English-speaking non-white	−1,648.080***	−1,206.114***	−1,488.347***
	(85.499)	(93.241)	(102.110)
Constant	515.932***	821.529***	359.583**
	(175.264)	(154.377)	(164.236)
Observations	5,053	5,548	7,153
R-squared	0.449	0.426	0.439

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from four to six rooms in a housing unit $((\log(6) - \log(4)) \cdot \beta_{i,1})$, holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table A3. Estimates of the willingness to pay for an older building.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	−3.252***	−4.675***	−2.664***
	(0.164)	(0.203)	(0.156)
College	−545.189***	−705.523***	−220.717***
	(36.352)	(47.113)	(33.670)
Famsize	−142.511***	−140.730***	−97.494***
	(13.478)	(17.228)	(12.997)
Age	−21.459***	−15.325***	−13.499***
	(1.663)	(2.158)	(1.530)
Married	−451.887***	−471.982***	−182.473***
	(41.694)	(52.459)	(39.153)
Male	−82.059**	−132.342***	43.730
	(33.984)	(42.978)	(31.577)
Spanish	552.782***	596.782***	−64.052
	(60.293)	(63.320)	(43.873)
Hindi and related	286.920***	76.473	−147.745
	(104.388)	(125.374)	(97.869)
Chinese	−400.119***	352.513**	131.751
	(143.370)	(171.568)	(130.022)
English-speaking non-white	754.405***	1,267.620***	83.121*
	(38.487)	(60.647)	(43.429)
Constant	−2,392.743***	−3,509.876***	−2,101.354***
	(78.894)	(100.412)	(69.852)
Observations	5,053	5,548	7,153
R-squared	0.399	0.353	0.134

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10 to 35 years in the age of a housing unit ((log(35) − log(10)) · $\beta_{i,2}$), holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table A4. Probit estimates of the demand for single detached houses.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	9.952***	11.23***	2.956***
	(0.798)	(0.824)	(0.42)
College	1025.801***	125.653	150.696*
	(140.648)	(154.09)	(84.366)
Famsize	1003.076***	1099.139***	274.881***
	(51.322)	(73.201)	(34.755)
Age	106.746***	61.614***	23.500***
	(6.594)	(7.185)	(3.907)
Married	2148.9***	1859.622***	576.255***
	(161.166)	(174.097)	(98.405)
Male	246.394*	585.02***	−26.401
	(127.764)	(140.055)	(78.856)
Spanish	−3157.475***	−2185.286***	−106.349
	(228.701)	(213.437)	(109.632)
Hindi and related	−2774.934***	−2690.828***	−136.263
	(388.068)	(414.83)	(247.22)
Chinese	−204.471	−3811.812***	−455.646
	(549.818)	(538.046)	(321.935)
English-speaking non-white	−2854.258***	−2806.349***	−276.519**
	(147.33)	(203.068)	(109.915)
Price of single detached unit	−1	−1	−1
Constant	−2997.641***	−1717.642***	−778.318***
	(328.627)	(393.534)	(190.542)
Pseudo-R2	0.425	0.411	0.101
Observations	5,053	5,548	7,153

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate probit estimate. The dependent variable is a dummy variable, "Single" equals 1 if the housing unit is single detached and 0 otherwise. *P*-values are reported in parentheses. The omitted category is female white English speakers who are single and have not graduated from college. Annual household income is measured in thousands of dollars, and the implicit price for single detached house is in dollars.

Table A5. Probit estimates of the demand for ownership.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	8.866***	7.873***	9.821***
	(0.798)	(0.583)	(0.797)
College	1762.106***	1050.408***	1823.097***
	(140.648)	(100.821)	(148.714)
Famsize	231.349***	164.193***	130.47**
	(51.322)	(37.62)	(57.139)
Age	63.020***	26.458***	56.343***
	(6.594)	(4.671)	(7.077)
Married	1540.468***	892.058***	1506.17***
	(161.166)	(113.78)	(173.705)
Male	303.647**	237.528***	161.916
	(127.764)	(91.787)	(136.388)
Spanish	−1955.127***	−118.354	−1042.531***
	(228.701)	(133.294)	(190.53)
Hindi and related	−1493.872***	−800.201***	−1096.082**
	(388.068)	(263.658)	(416.112)
Chinese	190.195	1982.427***	875.329
	(549.818)	(365.604)	(563.198)
English-speaking non-white	−2323.724***	−1034.745***	−1853.367***
	(147.33)	(131.638)	(195.448)
Price of ownership	−1	−1	−1
Constant	−170.800	−354.68	−574.561
	(328.627)	(225.602)	(369.757)
Pseudo-R2	0.396	0.266	0.230
Observations	5,053	5,548	7,153

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate probit estimate. The dependent variable is a dummy variable, "Own" equals 1 if the migrants own a home and 0 if they rent a home. *P*-values are reported in parentheses. The omitted category is female white English speakers who are single and have not graduated from college. Annual household income is measured in thousands of dollars, and the implicit price for a single detached house is in dollars.

Table A6. Estimates of the willingness to pay for a higher percentage of college educated households in the community.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	41.749***	43.954***	44.975***
	(1.878)	(1.589)	(1.883)
College	6,420.646***	5,917.598***	4,478.884***
	(417.165)	(369.591)	(405.706)
Famsize	1,960.873***	1,465.095***	2,778.890***
	(154.673)	(135.150)	(156.600)
Age	256.387***	156.794***	335.655***
	(19.086)	(16.930)	(18.438)
Married	5,405.325***	4,300.191***	9,077.404***
	(478.460)	(411.526)	(471.765)
Male	1,300.794***	1,058.294***	93.436
	(389.982)	(337.147)	(380.483)
Spanish	−8,906.960***	−4,114.947***	−9,574.449***
	(691.893)	(496.728)	(528.644)
Hindi and related	−4,836.778***	−2,745.227***	−7,673.110***
	(1,197.909)	(983.524)	(1,179.259)
Chinese	4,033.992**	−3,535.243***	−3,779.187**
	(1,645.251)	(1,345.902)	(1,566.690)
English-speaking non-white	−8,935.759***	−7,188.841***	−7,707.934***
	(441.661)	(475.758)	(523.291)
Constant	4,513.062***	8,781.568***	−77.853
	(905.358)	(787.706)	(841.676)
Observations	5,053	5,548	7,153
R-squared	0.439	0.393	0.432

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are college graduates $((\log(35\%) - \log(10\%)) \cdot \beta_{i,9})$, holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Annual household income is measured in thousands of dollars.

Table A7. Estimates of the willingness to pay for more English-Speaking Whites.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	24.162***	−24.557***	−15.245***
	(1.014)	(1.101)	(0.989)
College	2,813.012***	−2,110.578***	−2,019.040***
	(225.245)	(255.984)	(213.151)
Famsize	740.000***	−1,360.476***	−847.295***
	(83.514)	(93.607)	(82.275)
Age	87.367***	−101.729***	−126.318***
	(10.305)	(11.726)	(9.687)
Married	1,938.159***	−2,805.249***	−3,154.375***
	(258.340)	(285.029)	(247.857)
Male	641.202***	−903.011***	−322.511
	(210.568)	(233.513)	(199.899)
Spanish	−3,934.502***	4,659.829***	2,675.413***
	(373.582)	(344.041)	(277.740)
Hindi and related	−2,104.006***	3,055.582***	−2,377.695***
	(646.801)	(681.203)	(619.562)
Chinese	2,262.545**	3,832.952***	−925.317
	(888.339)	(932.191)	(823.112)
English-speaking non-white	−3,377.422***	5,800.150***	1,054.712***
	(238.471)	(329.517)	(274.928)
Constant	−2,241.780***	−2,226.147***	5,457.354***
	(488.840)	(545.577)	(442.202)
Observations	5,053	5,548	7,153
R-squared	0.353	0.349	0.245

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are white English speakers (that is, $WTP_{i,whiteEnglish}$), holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Both housing expenditure per year and annual household income are measured in thousands of dollars.

Table A8. Estimates of the willingness to pay for more Spanish speakers.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	0.598***	5.979***	18.563***
	(0.224)	(0.276)	(0.687)
College	44.564	559.404***	1,840.295***
	(49.785)	(64.087)	(147.978)
Famsize	293.639***	338.333***	1,011.669***
	(18.459)	(23.435)	(57.118)
Age	36.033***	32.481***	125.263***
	(2.278)	(2.936)	(6.725)
Married	666.096***	793.120***	3,316.319***
	(57.100)	(71.359)	(172.072)
Male	48.931	190.248***	45.930
	(46.541)	(58.462)	(138.777)
Spanish	−369.952***	−789.002***	−3,108.530***
	(82.572)	(86.133)	(192.818)
Hindi and related	−18.762	−631.121***	−2,141.112***
	(142.961)	(170.543)	(430.124)
Chinese	675.577***	−366.589	−1,368.369**
	(196.347)	(233.380)	(571.435)
English-speaking non-white	−450.904***	−815.180***	−2,547.028***
	(52.709)	(82.497)	(190.866)
Constant	−18.308	453.015***	−1,958.686***
	(108.047)	(136.588)	(306.994)
Observations	5,053	5,548	7,153
R-squared	0.241	0.329	0.449

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are white English speakers (that is, $WTP_{i,Spanish}$), holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Both housing expenditure per year and annual household income are measured in thousands of dollars.

Table A9. Estimates of the willingness to pay for more Hindi and related and related language speakers.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	−19.319***	2.480***	−6.285***
	(0.855)	(0.226)	(0.575)
College	−2,571.580***	3.115	−83.504
	(189.994)	(52.470)	(123.939)
Famsize	−809.103***	277.330***	−251.029***
	(70.444)	(19.187)	(47.840)
Age	−97.259***	20.273***	−13.379**
	(8.692)	(2.404)	(5.633)
Married	−2,188.116***	710.776***	−528.970***
	(217.910)	(58.424)	(144.119)
Male	−555.511***	139.008***	124.444
	(177.614)	(47.865)	(116.233)
Spanish	4,006.173***	−700.033***	184.307
	(315.117)	(70.520)	(161.495)
Hindi and related	2,051.493***	−547.779***	2,986.104***
	(545.577)	(139.630)	(360.250)
Chinese	−1,809.808**	−1,704.680***	1,590.744***
	(749.315)	(191.077)	(478.606)
English-speaking non-white	3,642.603***	−1,353.623***	1,141.516***
	(201.150)	(67.543)	(159.860)
Constant	2,226.035***	1,370.644***	−447.846*
	(412.338)	(111.830)	(257.123)
Observations	5,053	5,548	7,153
R-squared	0.405	0.267	0.066

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are white English speakers (that is, $WTP_{i,Hindi}$), holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Both housing expenditure per year and annual household income are measured in thousands of dollars.

Table A10. Estimates of the willingness to pay for More Chinese speakers.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	−9.040***	7.334***	13.799***
	(0.462)	(0.400)	(0.616)
College	−1,451.843***	496.052***	1,557.014***
	(102.686)	(93.054)	(132.627)
Famsize	−121.295***	466.311***	514.606***
	(38.073)	(34.027)	(51.193)
Age	−15.900***	33.020***	74.922***
	(4.698)	(4.263)	(6.028)
Married	−453.852***	794.003***	1,893.872***
	(117.774)	(103.613)	(154.222)
Male	−283.653***	313.165***	246.783**
	(95.995)	(84.886)	(124.381)
Spanish	1,599.988***	−1,654.707***	−1,829.079***
	(170.311)	(125.064)	(172.816)
Hindi and related	159.429	−1,334.592***	1,790.582***
	(294.868)	(247.628)	(385.504)
Chinese	−1,546.882***	−763.936**	1,271.226**
	(404.982)	(338.866)	(512.157)
English-speaking non-white	2,126.120***	−1,725.293***	−899.968***
	(108.716)	(119.785)	(171.066)
Constant	512.981**	−57.481	−2,890.728***
	(222.856)	(198.326)	(275.147)
Observations	5,053	5,548	7,153
R-squared	0.303	0.274	0.303

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are white English speakers (that is, $WTP_{i,Chinese}$), holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Both housing expenditure per year and annual household income are measured in thousands of dollars.

Table A11. Estimates of the willingness to pay for more English-Speaking nonwhites.

VARIABLES	Atlanta	Chicago	Dallas
Hhincome1000	1.722***	10.284***	−9.255***
	(0.340)	(0.448)	(0.446)
College	887.165***	1,188.717***	−1,083.698***
	(75.566)	(104.275)	(96.182)
Famsize	−145.803***	362.393***	−359.065***
	(28.018)	(38.131)	(37.126)
Age	−15.989***	22.939***	−49.645***
	(3.457)	(4.777)	(4.371)
Married	−110.554	711.239***	−1,261.855***
	(86.669)	(116.107)	(111.843)
Male	96.668	314.343***	−60.695
	(70.642)	(95.122)	(90.202)
Spanish	−957.048***	−1,759.454***	1,791.227***
	(125.330)	(140.145)	(125.327)
Hindi and related	31.366	−724.609***	56.917
	(216.991)	(277.489)	(279.571)
Chinese	177.875	−1,262.758***	−366.496
	(298.023)	(379.729)	(371.421)
English-speaking non-white	−1,550.863***	−2,317.273***	1,143.740***
	(80.003)	(134.229)	(124.059)
Constant	−384.865**	538.168**	−615.765***
	(163.998)	(222.241)	(199.539)
Observations	5,053	5,548	7,153
R-squared	0.157	0.309	0.285

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Each column presents a separate ordinary least squares regression. The dependent variable is a migrant's willingness to pay per year (in dollars) for an increase from 10% to 35% of community members who are white English speakers (that is, $WTP_{i,Englishnonwhite}$), holding all other housing attributes constant. The omitted category is female white English speakers who are single and have not graduated from college. Both housing expenditure per year and annual household income are measured in thousands of dollars.