Data Democratization and Spatial Heterogeneity in the Housing Market

RALPH MCLAUGHLIN
Veritas Urbis Economics

CHERYL YOUNG
Trulia

The internet has democratized data, improving access to information for millions of households worldwide and, in the process, making them better informed. The residential real estate industry is no exception, as the democratization of data has empowered the public with a wealth of information previously available only to real estate agents and other industry professionals. Left unexplored is how improved access to residential real estate data might affect spatial patterns of residential choice. In this paper, we posit that the democratization of real estate data ultimately makes the home search process more efficient. In turn, we argue, increased efficiency in the home search process has two potentially countervailing effects on the spatial heterogeneity of residential settlement patterns: (1) housing search choice sets expand to include properties in neighborhoods that are more diverse than choice sets developed by consumers in a less efficient environment, and (2) demand to live near amenities increases without an appropriate increase in housing supply, and thus "prices out" existing and future residents. However, we argue that the extent to which households might be priced out of a neighborhood is directly influenced not by data availability per se, but by the ease to which housing supply can be increased to meet demand in such locations. We therefore divide our policy recommendations into three specific efforts: (a) helping the market increase housing choice by reducing exclusionary and restrictive zoning policies in our nation's most expensive, amenity-laden markets, (b) giving housing choice voucher (HCV) recipients the option to hide their voucher status from land-lords during the application process, and (c) requiring a portion of the two-dollar per capita LIHTC funding to be used in Census tracts that qualify as high value.

This paper is structured as follows: in the next section, we present descriptive analysis of how consumers use online real estate data in the homebuying and rental search

process. Next, we review existing literature and develop a conceptual model of the effects of home search efficiency on spatial heterogeneity. In the penultimate section, we analyze measures of home value heterogeneity and racial diversity across the 100 largest housing markets in the United States. Last, we conclude with policy recommendations based on our findings in the previous section.

DEMOCRATIZATION OF DATA AND CONSUMER SEARCH BEHAVIOR IN REAL ESTATE

The internet has created, delivered, and disrupted the traditional process of finding, evaluating, and purchasing goods and services. The residential real estate industry is no exception, as both real estate and non-real estate-specific online purveyors of data have allowed homebuyers and renters to make more informed decisions.[1] Prior to the digital age, most real estate information was effectively proprietary: homebuyers and renters would have to contact real estate agents, landlords, and mortgage brokers to obtain the information needed for decision making. Sharing this information was a core service added to the traditional value chain by real estate professionals, and consumers would often start the homebuying process by gathering information with their help. Post-internet, the ubiquity of accessible data on homes, neighborhoods, and financial products has empowered consumers to gather these data on their own, by using either online real estate marketplaces (Redfin, Trulia, Zillow, etc.) or other internet services providing neighborhood-level information (Google Maps, Yelp, GreatSchools, etc.). Consumers today often gather much of this information before employing an agent, thereby altering the traditional real estate agent-consumer interaction.

While consumers now have access to an abundance of housing market data online, many continue to use real estate agents. For example, in a broad survey of 13,249 consumers by Zillow Group conducted in 2016, 87 percent of buyers use online resources and 75 percent select a real estate agent or broker to work with, suggesting that most buyers combine these two resources. In fact, buyers who utilize online resources are significantly more likely to also use an agent (77 percent versus 59 percent who do not use online resources).

Consumers are also initiating the home search process on their own. Only about half (51 percent) of buyers using an agent employ one at the beginning of their home search. Forty-one percent start the search process on their own, but ultimately use an agent before making an offer. This includes 27 percent who employ an agent after searching for a while but before touring homes, and an additional 14 percent who use an agent after visiting open houses but before submitting offers. Just 7 percent of all buyers who use an agent wait until they are ready to make an offer before enlisting the agent's services.[2] Below, we identify the types of online information that homebuyers and renters are likely using during their home search process.

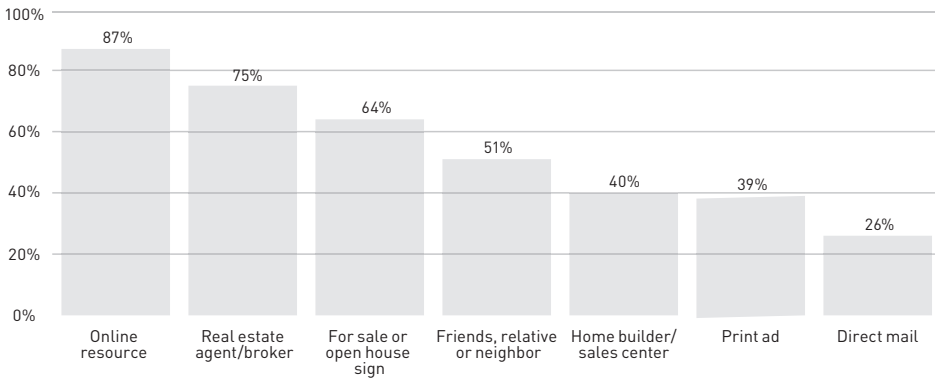**Figure 1: Top Resources Used to Search, Shop, or Purchase a Home**

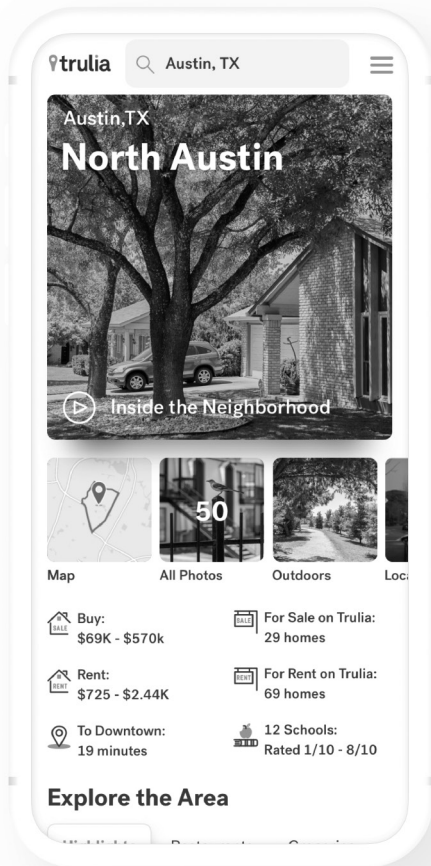**Figure 2: Property-Specific Information on Trulia**

Types of Data Available to Modern Homebuyers and Renters

In general, there are three primary types of online data available to homebuyers and renters. These are: (1) property-specific data, (2) neighborhood-specific data, and (3) user-specific data. We briefly describe and provide examples of each below.

Property-Specific Data. Detailed information about residential properties form the backbone of the most popular online real estate marketplaces, including consumer brands within Zillow Group (Zillow, Trulia, Hotpads, Streeteasy, NakedApartments), the Move brands (Realtor.com, Move.com, Doorsteps.com, SeniorHousingNet.com), Homes.com, and Redfin.com. These property-specific data include two subtypes: (a) public records data, which constitute the official record of home type, square footage, lot size, number of bedrooms and bathrooms, legal lot description, property taxes paid, year built, and sales history of a given

property, and (b) property listings data, which typically include listing price, pictures, seller/leasing description, agent contact information, and property details not available from public records. Public records originate from county assessor and recorder offices, while listing data can originate from a variety of sources, including real estate agents, multiple listing services (MLS), property owners (for-sale and for-rent), property managers, and third party aggregators. Three marketplaces — Zillow, Trulia and Redfin — also use these data to provide home value estimates.
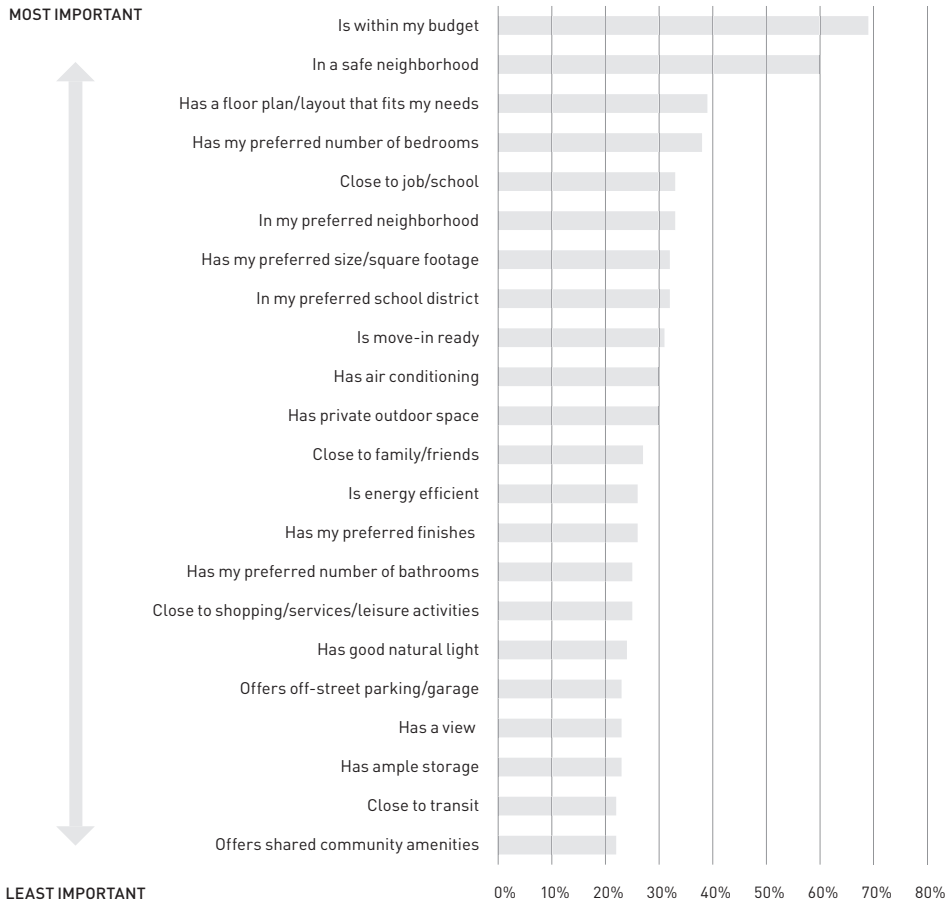
Figure 3: Neighborhood-Specific Data on Trulia

Neighborhood-Specific Data. Homebuyers and renters can also find a wealth of easily accessible information on neighborhoods surrounding properties of interest. Examples include local information on crime, school location and quality, business location and quality, natural hazards, and commute times. These data are available on many online real estate marketplaces, but often originate from third-party providers who also provide the information on their own websites. For example, Zillow, Trulia, Redfin, and Realtor.com all provide school rating information originating from GreatSchools.org, but the latter has its own site dedicated to providing the public with the same data. Other examples of neighborhood-specific data provided by online real estate marketplaces, but derived from third-party providers, include: the location and quality of local businesses provided by Yelp.com, crime data provided by SpotCrime.com and CrimeReports. com, and walking and transit scores provided by WalkScore.com.

Figure 4: Amenities of a Home - What Drives Home Selection

MOST IMPORTANT

Amenity	
Is within my budget	~68%
In a safe neighborhood	~60%
Has a floor plan/layout that fits my needs	~39%
Has my preferred number of bedrooms	~38%
Close to job/school	~32%
In my preferred neighborhood	~32%
Has my preferred size/square footage	~31%
In my preferred school district	~31%
Is move-in ready	~30%
Has air conditioning	~29%
Has private outdoor space	~29%
Close to family/friends	~26%
Is energy efficient	~25%
Has my preferred finishes	~25%
Has my preferred number of bathrooms	~24%
Close to shopping/services/leisure activities	~24%
Has good natural light	~23%
Offers off-street parking/garage	~22%
Has a view	~22%
Has ample storage	~22%
Close to transit	~21%
Offers shared community amenities	~21%

LEAST IMPORTANT

0%   10%   20%   30%   40%   50%   60%   70%   80%

User-Customized Data. In addition to property- and neighborhood-specific information, online real estate marketplaces also provide customized information for home-buyers and renters. Such information includes customized estimates of borrowing power, housing affordability, whether it is better to rent or buy a home, commute times, and mortgage quotes. This information is delivered primarily through stand-alone calculators on these websites, or in the case of commute time and affordability estimates, through individual property display pages. Similar to neighborhood-specific information, user-customized information can also be found on non-real estate websites. For example, commute times between two locations can be calculated using Google Maps, mortgage quotes can be obtained on Bankrate.com, and affordability estimates can be found on Nerdwallet.com.

Amenity Preferences

Using a survey of homebuyers conducted by Zillow Group, we find that consumers indeed value amenities that can be found using the three types of data discussed above. Figure 4 shows the share of households who value a specific amenity when selecting a home for more than 20 different amenity types. Online marketplaces offer data on most of these amenities. For example, nearly 70 percent of homebuyers say a home "within my budget" is important, which can be generated using online afford-ability calculators (user-customized data); 60 percent say a home in a "safe neighbor-hood" is important, and visualized crime data and other user-generated content offer a starting point (neighborhood-specific data); and nearly 40 percent say a home with "a floorplan/layout that fits my needs" is important, which can be found using online listings (property-specific data). Clearly, the democratization of real estate data online can help consumers more efficiently search for a home that meets their preferences.

However, the impacts of this democratization of data on household sorting and spatial heterogeneity are less clear. In the following section, we present two conceptual mechanisms by which increased search efficiency might affect household sorting.

NEIGHBORHOOD AMENITY PREFERENCES AND RESIDENTIAL CHOICE

In the home search process, traditional methods of information gathering — through word of mouth, scouring newspaper ads, or directly from a real estate agent — can result in information asymmetries and/or cost considerable time. Online real estate marketplaces complement these methods and introduce efficiency, particularly at the beginning of the process when homeseekers develop a choice set of properties. Theoretically, a more efficient search process should expand this choice set because it will allow more properties to be filtered in a given time period. Below, we discuss how spatial heterogeneity might be affected by a more efficient housing search process.

Household preferences have long played a role in spatial heterogeneity, as households tend to geographically sort themselves into communities with the bundle of goods and services best matching their preferences. This process of "voting with one's feet" is often based on local amenities, public goods, and tax rates.[3] This encourages people with like preferences to coalesce in groups that are able to take advantage of local collective benefits produced only through economies of scale.[4] Online search resources may enhance this self-sorting process.

Despite what is known about the role of preferences in housing choice, very little research has been done on the effects of online real estate marketplaces on traditional search methods. These methods have certainly changed in recent decades. In 1981, home searches were primarily initiated by consulting newspaper ads (22 percent) and relying on word of mouth (8 percent).[5] By 2016, online searches were the most

common method (44 percent), word of mouth dropped to fifth, and printed news-paper ads no longer registered. Online resources clearly provide more information at lower search costs; however, only emerging empirical evidence exists about how they have materially affected search patterns.

One of the few studies on the topic — a survey of house hunters in Wake County, North Carolina in the early 2000s — found that the use of the internet increased the number of homes visited by respondents in their search.[6] A more recent study shows that "search pressure," a measure of the online popularity of search locations, high-lights where demand for housing is high.[7] In these examples, online searches unveil the ability of consumers to conduct low-cost, extensive searches across geographies and property parameters to expand their housing choice sets.

Theoretically, democratization of real estate data could change neighborhood composi-tion in two countervailing ways. First, it may lead to more spatial heterogenity through the expansion of consumer housing choice sets. More competitive online shopping for mortgages, for example, could have such an effect: when households receive the lowest possible mortgage rate for their credit profile, their qualifying loan amount is maximized and thus so is their housing choice set. And improved search efficiency for properties could lead homeseekers to discover neighborhoods that they would not find through traditional methods, such as word of mouth or newspaper listings. As consumers filter a large volume of disparate information, biases towards certain neighborhoods might be lifted as they encounter homes that fit their preferences. The question in both cases is whether a larger choice set will also be qualitatively more diverse: if so, the expanded choice set would lead to greater diversity within cities, and vice versa.

Second, however, data democratization could lead to increased housing costs where certain amenities are in high demand, which, in turn, could exacerbate existing patterns of spatial homogeneity by pricing out lower-income households. Competition for amenities such as safe neighborhoods or good school districts could create an "amenity effect," raising the demand for a neighborhood to a level at which higher-income individuals could outbid others. Unless these neighborhoods had perfectly elastic supply, those that are outbid would be priced out. Exclusionary zoning is an existing practice that has created such amenity effects; it influences who settles in an area by increasing housing prices through either density restrictions or impact fees to cover public services.[8]

THE RELATIONSHIP BETWEEN RESIDENTIAL SEGREGATION AND HOUSING CHOICE

Both the passage of the Fair Housing Act in 1968 and the ability of online real estate marketplaces to reduce information asymmetries leading to discriminatory action has
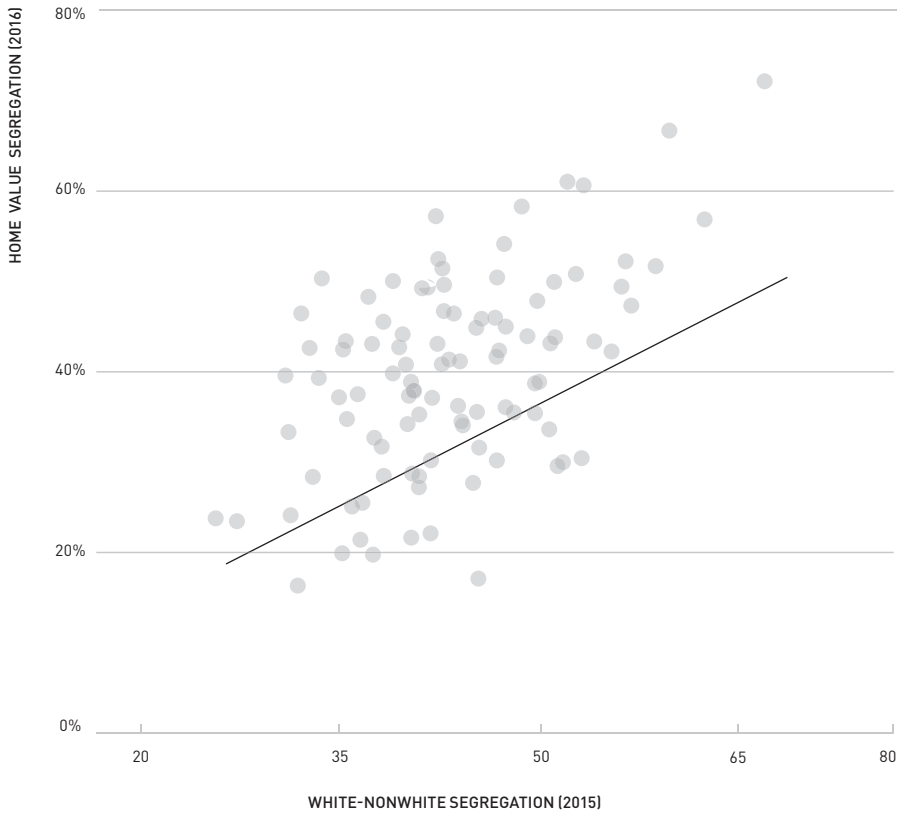
largely protected homebuyers and renters from explicit forms of discrimination. This section explores how historic processes of segregation led to path-dependent inertia in housing availability, particularly for those at the lower end of the housing market. Housing choice sets are constrained not just by the information available to housing consumers, but also by the diversity of housing supply across geographies.

The relative concentration of home values may help to explain persistent racial and income segregation in urban housing markets as homeseekers face expanding or contracting housing choices as they search across their neighborhoods of interest. These characteristics in the housing stock reflect ingrained effects of discrimination that continue to impact where and what type of housing is available. For example, the effects of redlining, which began in the late 1930s, can still be read in the socioeconomic outcomes of neighborhoods in cities today,[9] despite the fact that redlining is currently unlawful under the Fair Housing Act. More recently, in the wake of the Great Recession, researchers found that the incidence of foreclosures resulting from predatory lending of subprime mortgages was highly correlated with racial segregation.[10] The uneven impact of the foreclosure crisis on neighborhoods demonstrates that segregation can entrench inequalities of opportunity for years to come.

In order to examine whether racial segregation may be related to limited choice sets across metropolitan areas, we develop an index of home value segregation and compare it to measures of racial segregation. Racial segregation in this paper is measured by calculating the index of dissimilarity between white and nonwhite residents, or how unevenly distributed distinct groups are across Census tracts in a given metropolitan area.[11] Data on race are from the US Census's 2015 5-Year American Community Survey. Home value segregation is measured by calculating the ratio of the median value of homes in a given Census tract to the median value of all homes in the metropolitan area. These ratios are placed into one of six buckets, and the very low value and very high value Census tracts are added together to obtain the share of tracts in the most extreme value buckets and expressed through the home value segregation index.[12] This measure captures the share of Census tracts in a metropolitan area that contain the most extreme home values, and like measures of racial segregation, provides a snapshot of how unevenly distributed home values are across the metropolitan area. Home value data are provided by the online real estate marketplace Trulia. Racial segregation and home value segregation indices are calculated for the 100 metros[13] with the highest number of occupied housing units.

The link between measures of home value segregation and racial and income segregation is particularly revealing, and may help to explain persistent residential segregation. The relationship between the segregation of housing values and white-nonwhite segregation is positive and statistically significant, producing a Pearson's correlation

Figure 5: Home Value Segregation versus White-Nonwhite Segregation Across 100 Largest Metros

Source: US Census 2015 5-Year ACS and Trulia data.

coefficient of 0.53. Figure 5 shows that white-nonwhite segregation tends to rise as home value segregation rises across the most populated 100 metropolitan areas. In fact, variation in metropolitan segregation patterns can be at least partially explained by within-metro variation in how segregated housing choice sets are based on the polarization of home values neighborhood by neighborhood.

This relationship suggests that efforts to decrease racial and income segregation could be addressed by expanding choices in the housing market. The lack of housing diversity by price constrains homeseekers, particularly those that are lower-income. As a result, they are relegated to find homes where they are available, further entrenching racial and income segregation. Expanding housing choices entails providing equal access for households by reducing disparities in housing affordability. The next section concludes this paper and explores policy recommendations that expand housing choice despite the path-dependent inertia of housing supply.

SUMMARY AND POLICY RECOMMENDATIONS

In this paper, we have examined the various mechanisms by which the democratization of real estate data might affect spatial heterogeneity in the housing market. Specifically, we argue that democratization of real estate data improves home search efficiency. In turn, this improved efficiency can affect residential settlement patterns if (1) it encourages consumers to settle in neighborhoods that are systematically different than ones they would otherwise have settled in, and (2) demand to live near amenities increases without a proportionate increase in housing supply. Our findings also show that racial segregation is tied very closely to home value segregation in the largest US metropolitan areas, suggesting that increasing housing choice in the most expensive metropolitan submarkets might help reduce spatial patterns of racial homogeneity. Given the empirical evidence presented in the previous sections show that supply-side, rather than demand-side, factors are more to blame for patterns of residentia segregation in the U.s., we focus our policy recommendations on reform of existing housing choice policies rather than on household preferences.

Reform of Existing Housing Choice Policies

Our policy recommendations are subdivided into three specific efforts: (a) helping the market increase housing choice by reducing exclusionary and restrictive zoning policies in our nation's most expensive, amenity-laden markets; (b) giving housing choice voucher (HCV) recipients the option to hide their voucher status from landlords during the application process; and (c) requiring a portion of the two-dollar per capita LIHTC funding to be used in Census tracts that qualify as high value. We discuss these three in turn below.

Restrictive Zoning Policies. A large and growing body of scholarly work suggests that restrictive zoning laws make housing more expensive not only through supply restrictions but also because they reduce affordable housing choices.[14] While land use authority in the US is primarily in the hands of states and municipalities, the federal government can provide incentives for local and state governments to zone for more affordable housing types. Such incentives might include allocation of Department of Housing and Urban Development Community Development Block Grants, Department of Transportation Safety Grants, or Department of Education Local Education Agencies Grants that would be tied to increases in housing choice in areas with high demand and few affordable homes.

Since land use authority constitutionally rests with states, there is also room for addressing exlusionary policy reform at the state level. Such reforms would likely need to allocate both market-rate and below-market-rate housing unit targets for local governments. Enforcement measures are key to efficacy, and would need to take a balanced approach using both incentives and penalities, such as tying infrastructure

and school funding as well as local land use powers to these housing targets. Maryland's Priority Funding Areas and Florida's now defunct Growth Management Act provide relevant examples for how states might address the housing provision problems arising from exclusionary zoning.

Housing Choice Voucher Reform. Housing choice vouchers theoretically allow families to make geographically flexible housing decisions because the subsidy is attached to the person(s) rather than to a housing unit fixed in space. However, there have been obstacles to this flexibility in practice. Up until last year, fair market rent (FMR), which are used to determine the amount of voucher subsidies, was set at the metropolitan level. Given that rents are typically non-uniform across a housing market, the metropolitan-wide FMR would essentially limit HCV recipients to poorer neighborhoods. The Obama administration helped remedy this problem last year by introducing a limited roll-out of FMR at the zip code level, but implementation has since been postponed. Still, two challenges remain. First, landlords are not required to accept Section 8 tenants, and second, applicants are required to disclose their intent to use an HCV to landlords. Both requirements introduce avenues for discrimination based on landlords' willingness to accept HCVs. This could be remedied by making the Section 8 Housing Choice Voucher Program a single-blind process, where landlords are not allowed to ask about the voucher status of their applicants but voucher holders are free to use them for any property. Foe example, Seattle's fair housing laws provide a model for preventing landlords from discriminating against tenants based on their source of income and use of a Section 8 certificate or other subsidy programs.

Low-Income Housing Tax Credit Reform. Finally, place-based housing choice programs, such as Low-Income Housing Tax Credits (LIHTC), provide market-based subsidies to developers of affordable housing through issuing of resalable tax credits. Developers of affordable housing typically sell the credits they receive from the LIHTC program and use the proceeds to grow their equity stake in the project, which in turn improves their chances of attracting debt or equity financing to complete the project. To further increase the efficacy of the LIHTC program in expanding housing choice where choice is needed most, we propose requiring a proportion of the existing two-dollar per capita LIHTC funding to be used in the most income-segregated Census tracts. That proportion might be directly tied to current levels of spatial homogeneity in the housing market. For example, if the share of very high value Census tracts in an area were 20 percent, then 20 percent — or 40 cents — of the two-dollar per capita LIHTC allocation would need to be used in such Census tracts. In effect, this would allow municipalities with very high value tracts a choice in how to increase more affordable housing opportunities in their community: they could either amend their zoning laws to encourage lower-cost, market-rate housing that would remove

classification of the Census tract(s) as very high value, or allow LIHTC projects to be built using the spatial inequality-related LIHTCs.

Bibliography

Bayer, Patrick, Hanming Fang, and Robert McMillan. 2014. "Separate when Equal? Racial Inequality and Residential Segregation." *Journal of Urban Economics* 82: 32–48.

Boustan, Leah Platt. "Racial Residential Segregation in American Cities." 2012. In *The Oxford Handbook of Urban Economics and Planning*, edited by Nancy Brooks, Kieran Donaghy, and Gerrit-Jan Knaap, 318–39. Oxford: Oxford University Press.

Bratt, Rachel G., and Abigail Vladeck. 2014. "Addressing Restrictive Zoning for Affordable Housing: Experiences in Four States." *Housing Policy Debate* 24, no. 3: 594–636.

Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics* 123, no. 1: 177–218.

Cherif, Emna, and Delvin Grant. 2014. "Analysis of e-Business Models in Real Estate." *Electronic Commerce Research* 14, no. 1: 25–50.

Emerson, Michael O., Karen J. Chai, and George Yancey. 2001. "Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans." *American Sociological Review* 66, no. 6: 922–35.

Havekes, Esther, Michael Bader, and Maria Krysan. 2016. "Realizing Racial and Ethnic Neighborhood Preferences? Exploring the Mismatches Between What People Want, Where They Search, and Where They Live." *Population Research and Policy Review* 35, no. 1 (2016): 101–26.

Huffman, Forrest E., Arthur C. Nelson, Marc T. Smith, and Michael A. Stegman. 1988. "Who Bears the Burden of Development Impact Fees?" *Journal of the American Planning Association* 54, no. 1: 49–55.

Ihlanfeldt, Keith R., and Timothy M. Shaughnessy. 2004. "An Empirical Investigation of the Effects of Impact Fees on Housing and Land Markets." *Regional Science and Urban Economics* 34, no. 6: 639–61.

Kucuk, Umit S., and Sandeep Krishnamurthy. 2006. "An Analysis of Consumer Power on the Internet." *Technovation*, 27, no. 1–2: 47–56.

Levitt, Steven D., and Chad Syverson. 2008. "Market Distortions when Agents Are Better Informed: The Value of Information in Real Estate Transactions." *Review of Economics and Statistics* 90, no. 4: 599–611.

Madrigal, Alexis C. 2014. "The Racist Housing Policy that Made Your Neighborhood." *The Atlantic*, May 22.

Massey, Douglas S., and Jonathan Rothwell. 2009. "The Effect of Density Zoning on Racial Segregation in US Urban Areas." *Urban Affairs Review* 44, no. 6: 779–806.

McGuire, Martin. 1974. "Group Segregation and Optimal Jurisdictions." *Journal of Political Economy* 82, no. 1: 112–132.

National Association of Realtors. 2016. "Home Buyer and Seller Generational Trends Report 2016." Washington, D.C.: National Association of Realtors.

— — — . 2017. "Real Estate in a Digital Age 2017 Report." Washington, D.C.: National Association of Realtors.

Nelson, Robert K., LaDale Winling, Richard Marciano, Nathan Connolly, et al. 2015. "Mapping Inequality: Redlining in New Deal America." In *American Panorama: An Atlas of United States History*, edited by Robert K. Nelson and Edward L. Ayers. Web. https://dsl.richmond.edu/panorama/redlining (accessed December 15, 2016).

Palm, Risa, and Michelle A. Danis. 2001. "Residential Mobility: The Impacts of Web-based Information on the Search Process and Spatial Housing Choice Patterns." *Urban Geography* 22, no. 7: 641–55.

Piazzesi, Monika, Martin Schneider, and Johannes Stroebel. 2015. "Segmented Housing Search." Working Paper No. 20823. Cambridge, MA: National Bureau of Economic Research.

Rae, Alasdair. 2015. "Online Housing Search and the Geography of Submarkets." *Housing Studies* 30, no. 3: 453–72.

Rugh, Jacob S., and Douglas S. Massey. 2010. "Racial Segregation and the American Foreclosure Crisis." *American Sociological Review* 75, no. 5: 629–51.

Schelling, Thomas C. 1978. *Micromotives and Macrobehavior.* New York: Norton.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64, no. 5: 416–24.

Waldfogel, Joel. 2008. "The Median Voter and the Median Consumer: Local Private Goods and Population Composition." *Journal of Urban Economics* 63, no. 3: 567–82.

Zillow Group. 2016. "Consumer Housing Trends Report 2016."

Endnotes

1   Levitt and Syverson (2008).

2   Zillow Group (2016).

3   Tiebout (1956).

4   McGuire (1974).

5   National Association of Realtors (2017).

6   Palm and Danis (2001).

7   Rae (2015).

8   On density restrictions and racial segregation, see Massey and Rothwell (2009). On impact fees, see Huffman et al. (1988) and Ihlanfeldt and Shaughnessy (2004).

9   Madrigal (2014).

10   Rugh and Massey (2010).

11 White residents are those that report their race to be white alone and non-Hispanic. Nonwhite persons include people of all nonwhite races, plus people of Hispanic origin regardless of race.

12 The six value buckets, as defined by the ratio of median value in a Census tract to median value of the metro, are: 1) very low value — less than 0.67; 2) low value — 0.67 to 0.8; 3) low-middle value — 0.8 to 1.0; 4) high-middle value — 1.0 to 1.25; 5) high value — 1.25 to 1.5, and 6) very high value — greater than 1.5.

13 Metropolitan regions include Census Metropolitan Statistical Areas, or metropolitan divisions, where available.

14 Bratt and Vladeck (2014).