Title
Essays on Municipal Finance

Author
Howard, Troup

Publication Date
2020

Peer reviewed|Thesis/dissertation

Essays on Municipal Finance

by

Troup M. Howard

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Sraer, Chair
Professor Ulrike Malmendier
Professor Nancy Wallace
Professor Danny Yagan

Spring 2020

Essays on Municipal Finance

Abstract

Essays on Municipal Finance

by

Troup Howard

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor David Sraer, Chair

This dissertation focuses on the intersection of municipal finance and household finance. An average resident in the United States lives within the intersection of more than five sub-national government entities, each of which typically has autonomous taxing and budgetary authority. This work explores how fiscal choices made by these local governments affect the financial well-being of constituent households. The first half analyzes issues of equity in taxation and service provision, particularly with respect to local property taxes. The second half asks whether municipal capital structure choices, specifically relating to a particular form of public debt incurred by retirement systems, have a direct impact on household wealth.

The first chapter of my dissertation is coauthored with Carlos Avenancio-León. We document large, widespread racial and ethnic inequality in local property taxes. Using panel data covering 118 million homes in the United States, merged with geolocation detail for 75,000 taxing entities, we show a nationwide "assessment gap" which leads local governments to place a disproportionate fiscal burden on racial and ethnic minorities. We show that holding jurisdictions and property tax rates fixed, black and Hispanic residents nonetheless face a 10–13% higher tax burden for the same bundle of public services. For the median minority homeowner, this represents an additional cost of \$300–\$400 each year. At the 90th percentile of the national distribution, the excess tax burden is \$800 per year. The assessment gap arises through two channels. First, property assessments are less sensitive to neighborhood attributes than market prices are. This generates racially correlated spatial variation in tax burden within jurisdiction. Second, appeals behavior and appeals outcomes differ by race. This results in higher assessment growth rates for minority residents. We propose an alternate approach for constructing assessments based on small-geography home price indexes, and show that this would reduce inequality by at least 55–70%. This project provides insight into how institutional discrimination can arise and persist. In this setting, at the intersection of housing markets and government policy, overt racial discrimination has been illegal since the Fair Housing Act of 1968. We show that outcomes which are demonstrably not race-neutral can still arise in a setting where policies are explicitly constrained to be race-blind.

The second chapter considers real economic effects of public pension underfunding.

Most local governments in the United States sponsor a defined benefit pension plan for public employees. Unfunded public pension liabilities represent a shortfall between contractually protected commitments to future retirees and the assets held in trust to make these payments. As such, they are a particular form of public debt, ultimately backstopped by the taxpayer. State and local governments are subject to balanced-budget requirements, and therefore increases in unfunded pension liabilities imply, in expectation, the need to generate additional revenue or to reduce services at some future point. I test whether the increased expected costs represented by a shock to unfunded pension liabilities are capitalized into home prices. Using novel, hand-collected data on assets, liabilities, and fund flows for 200 of the largest county and municipal pension funds in the United States, I estimate whether increases in per-capita unfunded liabilities lower future house price growth. My measure of shifts in unfunded liabilities comes from large investment losses during the Great Recession. To address concerns that investment returns may be correlated with regional economic variables, I construct an instrument for fund investment returns from unexpected returns to broad asset classes. The identifying assumption is that the residual return for any broad asset class (extracted from a standard asset pricing model) is orthogonal to regional economic drivers. Using county-level measures of home price growth, I find little evidence of a strong link between pension obligations and home prices. I show that controlling for endogenous expenditure levels changes the sign of the estimated relationship, which is consistent with budgetary constraints implying a trade-off between spending on goods or services and payments to retirement systems. Using microdata on individual home transactions, I find a clear negative link between pension liabilities and home price growth, along with strong evidence that these effects are larger for more valuable properties. This suggests wealth-heterogeneity in how local residents weigh the benefits of public expenditures against the costs of public debt, and also shows that municipal financial structure is directly relevant to household financial well-being.

This dissertation is dedicated to my parents,
with love and deep gratitude for their endless support and encouragement.

Contents

List of Figures

List of Tables

Acknowledgments

Coauthor information: Chapter 1 of this dissertation is coauthored with Carlos Avenancio-León, Assistant Professor of Finance, Kelley School of Business at Indiana University.

Acknowledgments: I wish to thank first and foremost my committee chair, David Sraer, for being extraordinarily generous with his time, advice, insight, and encouragement during the past several years. I am so grateful for his mentorship during the program, which enormously influenced my intellectual growth along with the professional mark to which I aspire, both as a scholar and as a classroom instructor.

I have also benefited greatly from the wisdom and guidance of my other committee members: Ulrike Malmendier, Danny Yagan, and Nancy Wallace. Each of these four people has been a wonderful role model for conducting rigorous, relevant research in economics, and I plan to spend the next several decades trying to live up to their example.

So many other people deserve thanks and gratitude for helping me on this journey. For helpful comments and suggestions on my research, I am very grateful to Alan Auerbach, Hilary Hoynes, Amir Kermani, Pat Kline, Conrad Miller, Hoai-Luu Nguyen, Christopher Palmer, Jesse Rothstein, Emmanuel Saez, and Gabriel Zucman. Adair Morse offered consistent encouragement to an ex-English major tackling a Finance PhD, which was very meaningful during challenging times in the program. Gadi Barlevy, Arvind Krishnamurthy, Robert McDonald, and Brian Melzer inspired me to pursue a PhD in the first place, and surely bear much of the blame for my acceptance at Berkeley.

The PhD program is difficult, and were it not for the friendship of many colleagues, I doubt I would have arrived at the end. I have truly enjoyed discussing deep ideas, gleeful frivolities, and confusing equations with Maris Jensen, Marius Guenzel, Sanket Korgaonkar, Sheisha Kulkarni, Maria Kurakina, Chris Lako, Oren Reshef, Nick Sander, Aisling Scott, Jieyin Zeng, and Dayin Zhang, as well as many others. One of the luckiest things that happened to me at Berkeley was getting a desk next to Carlos Avenancio-León in my 1st year. His enthusiasm for research was overwhelming and infectious, and his willingness to discuss minutia of applied econometrics without any bound that I ever found.

I am also deeply grateful for love and support from friends and family outside of Berkeley. Clare Muñana taught me that education is transformative, and then never resented it when that led me to leave Chicago. John Aufdenkampe, Gretchen Kaluzny, Dave Sahota, and David Sobelsohn always offered laughs and sanity outside of academia. Steve Cicala has been providing good suggestions, great food, and a sympathetic ear for decades. And my parents, somehow, were always willing to pick up the telephone – for which I can never thank you enough.

Chapter 1

The Assessment Gap: Racial Inequalities in Property Taxation

1.1 Introduction

In the United States, the residential property tax is an ad valorem tax. The amount levied should be proportional to the value of the home. Authorizing legislation regularly makes explicit that the relevant concept of value is the market price of the property in a fair transaction. Property tax bills, however, are generated by applying the locally determined rate of taxation to an *assessed value*, which is a local official's projection of market price. Any wedge between market values and assessed values, therefore, generates some deviation from the intended rate of taxation. Equitable property tax administration requires the ratio of assessed value to market value to be the same for all residents within any particular taxing jurisdiction. This paper documents the existence of a widespread and large racial assessment gap: relative to market value, assessed values are significantly higher for minority residents. This assessment gap places a disproportionate fiscal burden on minority residents: within the same tax jurisdiction, black and Hispanic residents bear a 10–13% higher property tax burden than white residents.

We exploit a property-level dataset spanning most properties in the US, along with a comprehensive record of property transactions assembled from administrative data. We form assessment ratios by restricting the sample to homes for which we observe an assessment and a full-consideration, arm's-length sale within the same year. Using transacted prices ensures we construct the assessment ratio with an accurate measure of a home's fair market value. Because property taxes are levied by a wide range of local entities, which often only have partial geographic overlap, it is crucial to compare assessment ratios *within the same tax jurisdiction*. This also helps resolve a practical feature of property taxation, which is that assessments are rarely intended to be one-to-one with market value. The assessing entity chooses a scaling factor for assessments, which can range from less than 10% to 100%, and may change from one year to the next. As a result, variation in assessment ratios across tax jurisdictions may reflect either different levels of taxation or different scaling factors. To address this issue, we exploit a set of shapefiles that provide geographic delineation for the universe of local governments and other taxing entities in the U.S. We use these shapefiles to create unique taxing jurisdictions: properties belonging to the same jurisdiction face the same level of intended taxation, the same set of entities providing public services, and the same assessment scaling factor.

Our main empirical exercise compares assessment ratios – the ratio of a property's assessed value to its realized market value – within these tax jurisdictions. The average assessment ratio for a black resident in our sample is 12.7% higher than for a white resident. For black or Hispanic residents in aggregate, the average assessment gap is 9.8%. We show the assessment gap cannot be explained by racial or ethnic differences in realized market prices, nor is it simply a byproduct of racial wealth differences and the previously documented propensity for assessment ratios to be regressive (Baar, 1981, Black, 1977, Engle, 1975, McMillen and Weber, 2008, Paglin and Fogarty, 1972). As a result of the assessment gap, minority residents are therefore paying a significantly larger effective property tax rate for the same bundle of public services. For the median minority homeowner, the differential burden is an extra \$300–390 annually. This finding is strongly robust across most states in the U.S. We produce county-level estimates to characterize the distribution of this assessment gap. The average black homeowner in a county at the 90th percentile of the assessment gap distribution has a 27% higher assessment ratio, and would pay an extra \$790 annually in property tax.

We then explore two channels that drive these assessment gaps in the data. The first is spatial. We show that assessments are insufficiently sensitive to neighborhood-level attributes. Because of residential racial sorting, minority residents face, on average, different neighborhood characteristics than white residents (Ananat, 2011, Cutler et al., 1999, Massey and Denton, 1993). We show that assessed values and market prices align well on home-level characteristics, but diverge on tract-level attributes. In other words, market prices capitalize highly local factors but assessments are much less responsive. This generates spatial variation in the assessment ratio within jurisdiction. Residential spatial sorting leads this variation to correlate with homeowner race and ethnicity, generating just over half of the average assessment gap.

The second channel is a racial differential that persists even after conditioning away spatial factors. Within U.S. Census block groups, which represent regions of approximately 1,200 people, an average minority homeowner has an assessment 5–6% higher relative to market price than her nonminority neighbor. This latter finding is particularly surprising given that most assessors likely neither know, nor observe, homeowner race. We document that a significant portion of this effect arises from racial differentials in assessment appeals. To do so, we first analyze appeals in Cook County, the second largest county in the U.S. Using administrative court records, we show that minority homeowners: (i) are less likely to appeal their assessment, (ii) conditional on appealing, are also less likely to win, and (iii) conditional on success, typically receive a smaller reduction than nonminority residents. Then we show that national assessment patterns around changes in racial ownership are consistent with this channel: within the same property, assessment growth is significantly higher during the tenure of a black or Hispanic homeowner.

Finally, we propose a solution to at least partially address racial disparities in assessment gaps. This solution is tractable, and only uses publicly available data. As explained above, our results suggest that at least half of the racial disparity in assessment ratios emanates from the failure of assessments to account for geographic variations in neighborhood attributes. We describe an algorithm for generating assessments that relies on small-geography home price indexes. We show that simply linking assessment growth to zip-code-level indexes will reduce racial inequality by 55–70%. Racial inequality can be further reduced by using house price indexes that are more carefully calibrated to local geographies than zip code boundaries, which are well known to be drawn with little consideration for local characteristics.[1]

We believe the results we uncover in this paper represent a large source of racial inequality in the United States. Property taxes are directly relevant to nearly everyone in America. Many of the most salient public goods including education, policing, transportation infrastructure, and utilities are provided chiefly by local governments: cities, counties, towns, school districts, and other special purpose entities. For most of these local budgets, property tax revenue is the central financial pillar. For the average local government, property tax receipts comprise 56% of general revenue; and for the 14,000 independent public school districts in America, the average proportion is even higher at 74%.[2] The jurisdictions we form represent regions where residents have at least tacitly agreed upon some intended level of property taxation, and an associated level of public amenities provided with that revenue. Inequality within these jurisdictions, therefore, suggests that racial and ethnic minorities in the U.S. face different prices for the same set of public goods.

Much economic analysis of discrimination, including the canonical approaches of both Becker and Arrow/Phelps, focuses on how racial differences arise within a market environment. In contrast, our setting of local public finance allows us to study how racial differences in outcomes arise structurally, or institutionally, in a nonmarket environment. Since the Fair Housing Act of 1968, overt discrimination by race has been illegal. We show how inequalities arise nonetheless from institutional features of property tax administration. Our empirical analysis of assessment ratios is closely aligned to the legal concept of "disparate impact," a term that denotes group-level differences in outcomes between protected classes, one of which is race.[3] While differences are permitted, average differences between groups delineated by any protected class constitutes a discriminatory outcome, regardless of the process which generates the disparity.[4] Our results show that race-blind

[1]Stevens (2006) summarizes historical details in a Congressional Research Services report.

[2]General revenue excludes state and federal transfers, which can be large. General revenue is the funding stream which the local entity can direct affect.

[3]The others are: "religion, sex, handicap, familial status, or national origin" (42 USC 3604–3605). In 2015, the U.S. Supreme Court affirmed that disparate impact is the standard by which to legally evaluate discrimination claims in the housing market. *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc* (135 S. Ct. 2507). This remains an evolving area of jurisprudence. In August 2019 the Department of Housing and Urban Development issued a call for comments on a new proposed rule which would change the legal standards for establishing a disparate impact claim; this proposal is still outstanding as of this writing.

[4]An exception is if the process is "necessary to achieve one or more substantial, legitimate, nondiscriminatory interests" of the government (Atuahene, 2017).

policies may still generate outcomes which are not race-neutral.

We make four main contributions to the literature. First, we contribute to the literature on racial disparities in the property tax. There is a long history of activism seeking to address racially motivated over-assessment of residential property. Kahrl (2016) describes property tax rates as central to African American political mobilization during the Reconstruction era, and also provides examples of homeowners in the 1920s and 1930s suing local governments for relief from discriminatory assessments. Rothstein (2017) details the same concerns arising in the 1960s and 1970s. Atuahene and Berry (2019) estimate a causal link between inflated assessments and tax foreclosures within one county in Michigan between 2009 and 2015.[5] We build upon this research by: (i) documenting the widespread, contemporaneous presence of assessment gaps using comprehensive national data; (ii) providing a more refined notion of the proper taxing jurisdiction to precisely quantify the breadth and magnitude of disparate impact in property tax burden in the post-civil rights era; (iii) using administrative data to link individual properties with homeowner race and ethnicity rather than relying on regional demographic aggregates; and (iv) evaluating the mechanism through which the assessment gap arises.

Second, we show a new mechanism which can help explain the large and persistent black-white wealth gap. One strand of the broad literature studying racial inequalities in wealth has focused on the role of geography and spatial sorting (Cutler et al., 1999, Gittleman and Wolff, 2004, Card and Rothstein, 2007, Charles and Guryan, 2008, Ananat, 2011, Chetty et al., 2014, 2018). We show a new source of racial inequality in wealth, which operates through a public finance channel and is largely generated by spatial sorting within taxing jurisdiction. The effect is highly persistent (property taxes must be paid every year) and exists in most locations across the county. At the median, the assessment gap results in a black homeowner paying approximately $390 dollars more each year. This is a very large number, given that median black household net worth is $13,000, of which only $4,000 is in liquid assets.[6] For any discount rate below 3%, the stream of incremental tax payments suggested by our findings represents an excess tax burden which exceeds total household wealth for the median black family. Not only does this inhibit wealth building directly, it may well distort home ownership and financing choices for minority residents, further exacerbating the wealth gap. Much scholarship on racial wealth inequality focuses on channels that affect education and wages; here we show a channel which operates on wealth directly.

Third, we contribute to a small but growing literature that explores the bias and distributional consequences of algorithms and statistical procedures. An active debate in this literature is whether using race or racially correlated variables will reduce or exacerbate bias in outcomes. Bartlett et al. (2018) show that FinTech algorithms in the mortgage market generate higher interest rates for Hispanic and African American borrowers, although rejections are lower relative to face-to-face lending. Fuster et al. (2018) show that black and Hispanic borrowers are less likely to gain from increased precision in credit pre-

[5]In a related article Atuahene (2017) argues that present-day assessment practices in the city of Detroit should be considered federally illegal under the Fair Housing Act.

[6]Survey of Income and Program Participation (2016).

diction generated by machine-learning models. Kleinberg et al. (2018) argue that allowing algorithms to use protected class variables directly will provide an effective mechanism for reducing bias in decision making. Our results in this setting support the latter notion. Automated valuation and mass appraisal is an algorithmic prediction problem. We show that assessments will more closely track market values if the demographic composition of local areas is considered, simply because this variable is a strong statistical proxy for many factors that influence market prices.

Finally, we add to an extensive literature examining racial differences across a diverse range of outcomes including health (Schulman et al., 1999, Williams, 2012), employment (Donohue III and Heckman, 1991, Card and Lemieux, 1996, Bertrand and Mullainathan, 2004), criminal justice (Knowles et al., 2001, Arnold et al., 2018), and residential housing markets (Charles and Hurst, 2002, Bayer et al., 2007, Card et al., 2008, Bayer et al., 2017). That U.S. government policies in the early half of the twentieth century deliberately promoted racial segregation and discrimination in housing markets has been widely documented (Rothstein, 2017); however overt discrimination by race has been illegal since the 1960s. We use modern data to show that minority homeowners still face financial discrimination generated at the intersection of housing markets and local public institutions.

The paper proceeds as follows. Section 1.2 describes the typical structure of local property taxation and highlights important institutional details that pose econometric challenges. Section 1.3 outlines our empirical strategy. Section 1.4 details the data sets we use. Section 1.5 presents the results. Section 1.6 outlines potential policy approaches for achieving a equitable tax burden. Section 1.7 concludes.

1.2 Setting and Institutional Detail

1.2.1 Local Governments

Property taxation in the United States is chiefly a feature of local governments. Government authority in the United States is organized at three levels: federal, state and local.[7] These levels are roughly hierarchical. State constitutions and laws empower local units of governments, while retaining preeminence in the case of any regulatory conflict. Local units are empowered either by an explicit enumeration of powers,[8] or through "home rule" provisions which grant local units all authority not explicitly reserved for the state.[9] Counties and cities are the most prominent example of local governments, though as discussed below there are many other relevant types of local entities.

Although lowest in the hierarchy, state and local governments tend to have the most salient day-to-day impact on the lives of residents. Schooling, public safety (police and

[7]The intent of this overview is to help orient any reader unfamiliar with the general structure of American government. It is very much not a careful description of American federalism or the ways in which authority is mediated between levels of government.

[8]Oklahoma is one example. See Article XVIII–1 of the State Constitution and the extensive codification of authority in Title 11 of the Oklahoma statues.

[9]Montana is one example. Typical language appears in Montana Code Annotated 2019, Title 7, Ch. 1, Part 1.101, and Article XI, Part XI, Section 6 of the state Constitution.

fire), infrastructure, and transportation are all amenities that are chiefly provided by local governments with varying degrees of state and federal support. The vast majority of local government units impose a property tax, and these revenues are the central fiscal pillar of local government budgets.[10] They are the largest source other than intergovernmental transfers, and are very stable year-over-year (see Table 1.1). For some important local amenities, property taxes are even more crucial: independent school districts on average generate 74% of their revenue from the property tax. In 2012, local units of governments collected an aggregate of $433B in property taxes.[11]

Local units have broad discretion to set the level of intended tax burden. There is wide regional heterogeneity in the mechanism used to change the tax rate. Two approaches are most common: either voters have direct input into property tax levels at the ballot box, or they delegate this authority to locally elected officials (who may or may not redelegate this authority to appointed individuals). Often, the intended tax burden is implicitly defined: a certain level of spending will be approved (through either of the previously mentioned mechanisms) and then this amount will be divided by the total value of the local property, yielding an implicit tax rate. For this analysis, what is important is that local units set their own intended level of taxation each year.[12]

1.2.2 Effective Rates Depend on Assessments

The residential property tax is implemented as an ad valorem tax: residents are taxed some proportion of the value of their property.[13] This concept is regularly explicitly delineated in American law. Virtually every state has language in its constitution or legislative code carefully specifying that property taxation is intended to represent a proportional burden on the fair market value of the real property.[14]

From the starting point of a pure ad-valorem tax, most localities provide for deliberate adjustment in the form of property tax exemptions. Based on certain eligibility criteria, a homeowner is shielded from having to pay taxes on some portion of the home's value. In Florida, for example, homeowners are exempt from property taxation on the first $25,000 of home value, but only if that home is their primary residence.[15] Another common exemption applies only to senior citizens. Because eligibility varies by resident within a region,

[10] Some states impose a state-wide property tax levy, but the major source of state revenues are sales and income taxes. In 2012, state governments accounted for only 3% of all property taxes raised; local units comprised the remainder.

[11] Authors' calculations using Census of Governments data; figure given is nominal dollars.

[12] There are often legislative constraints that limit the rate of annual change.

[13] While there are examples of localities imposing fixed, or unit property taxes, these tend to be specific levies approved to fund a particular project (or to cover debt service for a given bond issuance). We do not have any way of providing an aggregate breakdown of tax dollars raised by ad valorem taxation versus unit taxes; in every region we have looked at specifically, unit taxation is a very small portion of overall proceeds.

[14] One example from Georgia: "'Fair market value of property' means the amount a knowledgeable buyer would pay for the property and a willing seller would accept for the property at an arm's length, bona fide sale." 2018 Georgia Code, Title 48, Chapter 5, Article 1(3).

[15] 2019 Florida Statues 196.031.1(a).

property tax exemptions on the whole will induce variation in effective tax rates within a region where the ad-valorem rate of taxation is held constant. Our empirical strategy is carefully designed to focus *only* on inequality in the purely ad-valorem component of the property tax: the component which by law should be the same for everyone. We elaborate on this in Section 1.3 and Section 1.5.5.

As described in Section 1.2.1, for every taxing government, a tax rate exists (either explicitly or implicitly defined). We refer to this as a "policy rate" to highlight that this rate is a political or legislative lever which can be adjusted to change the desired tax burden. However, it is crucial to realize that this policy rate is, alone, not sufficient to characterize the effective tax rate. This is because tax bills are calculated by applying the policy rate to an "assessment," which is a local official's projection of a home's value. For every home in America, there is some bureaucratic entity charged with producing an assessed value for that property. Very often this responsibility lies with county governments, and is executed through a county assessor's office. In some regions, the authority devolves to the township level.[16] These property assessments are a legal determination of value for purposes of the taxing entity, and will be a central object of our analysis. Every home in a taxing jurisdiction is assigned an assessment for each tax year. These assessments may be revised or reestimated annually, biannually, or in some regions even less often.[17] We observe realized assessments for all homes in our dataset annually.

If the policy rate is 5% and the home's assessed value is $100,000, then the homeowner will receive a tax bill of $5,000: a 5% tax applied to the $100,000 assessment. However, and perhaps surprisingly, nothing in the previous sentence necessarily implies that the market value of the home is $100,000. While the natural intuition might be to assume that assessed values track market values one-to-one, this is not the case for most of the country. Local units have a free scaling parameter in choosing how to produce assessments. States may mandate a particular level: Alabama specifies that residential assessments should be 10% of market value.[18] Thus, if the home described in the beginning of this paragraph were in Alabama, the $100,000 assessment would, in fact, imply a market value of $1,000,000.[19] The effective rate of taxation for this home would be 0.5%: the homeowner pays $5,000 in tax on a million-dollar asset.

In absence of state regulation, local units choose their own scaling factor. Sometimes local practices conflict with state targets, adding another layer of administrative complexity: Illinois state dictates that assessments should be 33%, but Cook County within Illinois uses 10% as an assessment target. To reconcile these, Illinois state law mandates an adjustment be applied to local assessments in order to achieve the state-level target.[20] Figure 1.2 shows the raw distribution of realized scaling factors in our data. As this figure shows, there are jurisdictions which appear to be targeting 100%, but there are also many jurisdictions

[16] This is relatively more common in the New England states, for example.

[17] Cook County, IL for example, conducts assessments on a triannual scale; each property is assessed every third year.

[18] Code of Alabama, Section 40, Chapter 8, Section 1.

[19] Errors in assessed values are central to this paper. This calculation assumes an accurate assessment for purposes of the example.

[20] Illinois Compiled Statutes: 35 ILCS 200/Art. 17.

which are clearly targeting another number.

Economically speaking, assessment scaling factors are meaningless. It is simply a free choice parameter for the local government. Consider tax revenue in jurisdiction j and year t as a function of two variables: $revenue_{jt} = f(rate_{jt}, scaling_{jt})$. Revenue is stable if the rate is doubled when the scaling factor is halved. However, as a consequence of the regional heterogeneity in scaling, the econometrician observing only two homes, each assessed at $50,000, can make no inference about relative market value of these properties and hence also cannot discern whether the tax burden for these properties should be the same.

We do not observe scaling factors. In order to draw meaningful inference from variation in assessment ratios, we must therefore be able to discern how all properties map geographically to governments. The following section outlines the challenges this poses, and describes our solution, which is to form "taxing jurisdictions" that hold fixed the choice of scaling factor, as well as the level of intended taxation.

1.2.3 Forming Taxing Jurisdictions

Local governments are highly spatially complex. Across the U.S. more than 75,000 entities potentially impose a property tax. Homeowners typically face taxation from multiple local units simultaneously. As mentioned, cities and counties are key examples of local government units. However, it is very common for regions to have a range of separate autonomous taxing entities. Chief examples here are: school districts, park districts, and municipal utility districts. Taxing authority may also be embedded in a special purpose district like an airport authority or regional economic development initiative. As a rule, the boundaries of these units are not naturally coincident. Counties are a complete partition of space in the US: every point in a given state lies in exactly one county. However, no such logical precision applies to other local entities. Cities often lie across county boundaries. In low-population-density areas, school districts often cover multiple towns (and potentially portions of different counties); in urban areas, there may be multiple school districts within a given metropolitan region. Units like park districts or utility districts typically have a delineation governed by a service area that reflects physical geography and may have little to do with nearby civic boundaries. Excluding state governments, the average home in the United States is touched by 4.5 local entities, all of which potentially levy a property tax.[21]

Our empirical goal is to explore whether minority residents face a higher tax burden than their white neighbors, *conditional on holding intended taxation fixed*. We are not asking whether minority residents tend to live in regions with more (or fewer) public services, which would naturally suggest higher (or lower) taxes. Rather, we wish to compare two residents for whom the tax burden should be identical: served by the same set of governments, receiving the same bundle of public goods, and facing the same policy tax rate. This analysis is only possible because we find a method for discerning the networks of overlapping governments that touch properties in our sample. We accomplish this by mapping the geolocation of property parcels onto geographic shapefiles for the universe of local governments in the United States. Colloquially, we will use the term "jurisdiction" to

[21] Author's calculations using Atlas Muni Data shapefiles.

denote a geographic region served by a unique set of taxing entities. Every home within a given jurisdiction faces the same set of governments providing services and imposing property taxes. Every piece of empirical evidence in this paper rests on holding these unique government networks fixed.

An important nuance warrants additional discussion. Every resident in a jurisdiction receives public goods and services from the same set of taxing authorities. This does *not*, however, mean that the quality of services received is equal. Consider public education: the taxing entity would be a school district.[22] However, districts may have tens or hundreds of schools, and certainly the quality of education delivered may vary from school building to school building. In the example of a school district, the implicit social contract underlying the property tax is: homeowners agree to pay X% of the value of their home in exchange for the right to receive education services. More generally, any homebuyer implicitly accepts the bargain to make some level of (ad-valorem) tax payment to a set of governmental entities in exchange for a bundle of goods and services. This paper documents inequality on the payment side of this social bargain. Our jurisdictions are formed to create regions where the legal commitment to pay (by virtue of homeownership) is the same for everyone. Of course it is possible that there is inequality on the service provision side as well. Our estimates will not capture any inequality relating to differential service provision within agency.

Panel A of Figure 1.3 illustrates our approach in a stylized example. There are three governments in this example: the county, which contains a city and an independent school district. The city and the school district have partial overlap. This spatial overlay of governments generates 4 taxing jurisdictions. Jurisdiction one contains those homes which receive services from, and are taxed by, the county alone. Homes in jurisdiction two are served and taxed by both the county and the city. Homes in jurisdiction three are served and taxed by all governments, and homes in jurisdiction four are served and taxed by the school district and the county. Panel B of Figure 1.3 highlights our focus on within-jurisdiction inequality. In this stylized example, the county realizes assessment ratios of either 50% or 20%. This generates inequality in the taxing jurisdiction comprised of just the county: there is large (binary) variation in assessment ratio. This does not generate inequality in the jurisdiction served by both the city and the county: everyone paying taxes and receiving public services in this region has the same assessment ratio. For any cross-jurisdiction comparisons, we cannot rule out Tiebout sorting along preferences for public goods or intended levels of property tax. Our focus is solely on inequality between residents who are subject to the same set of taxes and who have access to the same bundle of public goods.

The example in Panel A of Figure 1.3 is, in fact, quite common across the county. However, jurisdictions can be complex, especially in more urban regions. Figure 1.4 shows the example of Harris County, Texas. Including the county, there are 12 local units of government which overlap in varying combinations. Each combination forms a distinct

[22] In practice, there is a distinction between independent school districts which have autonomous taxing authority and school districts which are subordinate to, and funded by, an upstream government (typically a city or county). While our taxing jurisdictions carefully address this distinction, the difference is not important for the discussion of heterogeneity in service provision.

jurisdiction. One such jurisdiction is the region defined by the nexus of all 12 governments (this region is not visually identifiable in Figure 1.4). In our full sample, we observe a market transaction (paired with an assessment) for approximately 100 homes within this particular jurisdiction. This is a relatively small jurisdiction. Others are the size of cities and encompass tens of thousands of home transactions.[23]

The within-jurisdiction analysis is crucial in two distinct ways: (i) it holds fixed the level of intended tax burden, and (ii) it holds fixed the regionally chosen scaling factor. As we will describe and support, equity in the property tax demands that assessment ratios be constant within jurisdiction. This is not a controversial notion: it is often mandated in state legal codes,[24] and is also a primary tenet of best-practice standards for professional assessors.[25]

Assessment ratios, therefore, are only relevant because they are a sufficient statistic for inequality in effective tax rates. This logical relationship only holds, however, within a region where everyone should be facing the same tax burden. No meaningful comparison in tax burden is possible if we compare residents who pay taxes to (and thereby receive services from) a different set of governments. The way we form jurisdictions ensures that we only compare tax burden between residents paying a tax to the same set of governments.

Similarly, we cannot meaningfully compare assessment ratios between two homeowners who live in regions which are simply targeting a different assessment ratio. Our data does not reflect which entity produces assessed values, nor the target assessment ratio. Conducting our analysis within jurisdiction – according to our precise notion thereof – ensures that no error is introduced by an inability to observe local unit heterogeneity in assessment practices.

While our jurisdictions have both a natural economic and political interpretation, it is certainly reasonable to wonder whether our results are driven in any way by the partitioning of geography. We can test this fairly directly. Practically speaking, assessments are most commonly done at the county level. Often this is a provision of state law, but even when not required, it seems that either custom or natural considerations of efficiency and resource management often result in counties "owning" assessments. While it does not make any sense to compare effective tax rates within county (because so many sub-county units impose other property taxes and provide services), if target assessment ratios are unlikely to vary within county, we can meaningfully compare assessment ratios within county instead of within jurisdiction. In Section 1.5, we show that our baseline results establishing racial differences in assessment ratio are robust to conducting our analysis within county. Within county estimations, in fact, generate slightly higher estimates. Our preferred specifications all use the more rigorous partitioning into jurisdictions of unique overlapping governments.

[23]In some regions, all substate units of government are spatially aligned; Philadelphia is one such example: the county and city of Philadelphia, along with the school system, are all entirely coincident. This is relatively rare.

[24]See for example, Michigan Compiled Laws, Section 211.34(2).

[25]For example: *Guidance on International Mass Appraisal and Related Tax Policy*, International Association of Assessing Officers, 2014.

1.3 Empirical Strategy

The outline of our approach is as follows. We first define a notion of equitable tax administration within a jurisdiction. Then we show that within-jurisdiction variation in assessment ratios is an empirical sufficient statistic for rejecting the equitable tax null.

Our notion of equity relies on the ad valorem nature of the property tax, and the fact that taxes are levied on assessed values. We consider first a property tax system which does not establish individual tax exemptions, and then show the theory easily incorporates an arbitrary exemption structure. Let i denote property, j taxing jurisdiction, and t year. Further, let V^* be the true value of the property being taxed. Given a tax rate of r_{jt}^{eff}, by definition an equitable ad valorem tax must satisfy:

$$equitable\ tax_{ijt} = r_{jt}^{eff} V_{ijt}^*. \tag{1.1}$$

Note that r^{eff} is an effective tax rate. Let c be the local scaling factor for assessments, and let r be the policy rate that rationalizes equation 1.1: $r_{jt}^{eff} = r_{jt}c_{jt}$. This last equation simply reflects that if assessments are scaled to be half of market value, the policy rate must double in order to achieve the level of tax burden implied by r^{eff}.

Property tax bills are actually generated by applying this policy rate to an assessed valuation, A_{ijt}:

$$actual\ tax_{ijt} = r_{jt}A_{ijt}. \tag{1.2}$$

Our equitable tax null is simply that $actual\ tax_{ijt} = equitable\ tax_{ijt}$. We observe A_{ijt}, the realized assessed valuation assigned to the house. We observe market prices for homes, M_{ijt}, and accordingly will let $M_{ijt} = V_{ijt}^*$.[26] Equating 1.1 and 1.2, and taking logs yields a necessary condition for equitable administration of an ad valorem tax:

$$ln(A_{ijt}) - ln(M_{ijt}) = ln(c_{jt}) := \gamma_{jt} \qquad \forall i. \tag{1.3}$$

Equation 1.3 is a theoretical statement that does not allow any errors at all in assessments.

Empirically, we define a deviation from our fair tax benchmark in context of arbitrary demographic delineations. Partition the homes of any jurisdiction into M subsets based on any demographic characteristic, and denote by $m \in \{1, 2...M\}$. Let $\bar{c}_{mjt} := \frac{1}{N}\sum_{i\in m} c_{ijt}$. Our fair taxation null is:

$$\bar{c}_{mjt} = \bar{c}_{m'jt} \qquad \forall m, m'. \tag{1.4}$$

Equation 1.3 says that assessment ratios should not vary at all within jurisdiction. While strictly true, this represents unattainable precision. Equation 1.4 says that average assessment ratios should not vary within jurisdiction by demographic group. For an ad valorem tax burden within a jurisdiction, taxes levied should be a constant proportion of market value. However, property taxes are calculated as a proportion of assessed value. Thus, an increase of assessments relative to market value represents an increase in the overall tax

[26] It is worth reiterating that state laws explicitly state that property taxation should be levied upon the "fair cash value" that would be received in an arm's-length transaction. Therefore, our reliance on market prices is not a strong statement about perfect markets or market efficiency, but rather a reflection of the legal intent underlying the taxation.

burden. Correspondingly, if group m has higher assessed valuations relative to market than group m', then group m faces a higher tax burden.

We test inequality by racial and ethnic group with estimating equation:

$$ln(A_{ijt}) - ln(M_{ijt}) = \gamma_{jt} + \beta^r race_{ijt} + \epsilon_{ijt}. \tag{1.5}$$

Here *race* is a vector of indicator variables for racial and ethnic groups. The formulation of equation 1.5 is motivated by the legal notion of disparate impact. The Department of Housing and Urban Development states: "[a] practice has a discriminatory effect where it actually or predictably results in a disparate impact on a group of persons[...] because of race, color, religion, sex, handicap, familial status, or national origin."[27] As the left-hand side of equation 1.5 is a sufficient statistic for within-jurisdiction tax burden, this formulation is of primary interest for establishing inequality by race. The fixed effect γ_{jt} absorbs the realized average assessment ratio within jurisdiction. Then, since *race* is a categorical variable, β^r is a vector of estimated group-level deviations from average realized assessment ratio. If β^W, the average assessment ratio for white residents, is statistically different from β^M, the average assessment ratio for any grouping of minority residents, this would be evidence of inequality in tax burden.

The derivation above abstracts away from tax exemptions. As noted in Section 1.2.2, most jurisdictions establish individual-level criteria for tax exemptions. Incorporating this, the expressions for equitable tax and actual tax bills become:

$$actual\ tax_{ijt} = r_{jt}(A_{ijt} - E_{jt}(i)) \tag{1.6}$$

$$equitable\ tax_{ijt} = r_{jt}^{eff}(V_{ijt}^{*} - E_{jt}^{*}(i)). \tag{1.7}$$

$E_{jt}(i)$ is the homeowner-level exemption established by law, and is written as a function of i to highlight that this depends on personal characteristics (e.g. age or residency status). $E_{jt}^{*}(i)$ is the corresponding portion of the market value that is shielded by tax. This differs from E_{jt} only due to the scaling factor c_{jt}. If assessments in a given jurisdiction are done at 50% of market value, an exemption that reduces assessed value by \$10,000 corresponds to a reduction in market value of \$20,000: $E_{jt}^{*} = c_{jt}E_{jt}$. Given this relationship, the equitable tax benchmark implied by equations 1.6 and 1.7 is equivalent to equation 1.3.

Finally, before proceeding, we highlight an econometric point. Given the way tax bills are generated, our equitable benchmark implies a constant simple assessment ratio. We use logged values in our specifications because of heterogeneity in scaling factors. As we show in Figure 1.2, some regions target an assessment ratio below 10%; others 100%, with a wide range in between. Therefore an aggregate regression using simple ratios, which implies additive rather than proportional deviations around the mean ratio, does not have a natural interpretation. Scale invariance makes log differences a natural solution; however, this does generate a Jensen's inequality term. If minority residents select into homes with higher within-jurisdiction price variation than white residents, this would bias our estimates upwards. In Section 1.5, we show that our results are robust to running regressions on simple ratios rather than log differences.

[27] 24 CFR 100.500(a).

1.4 Data

The core research design of this paper rests on combining data from three sources: 1) annual property-level records of assessments, transactions, home characteristics and geolocation from ATTOM, 2) Geographic Information System (GIS) detail on local government boundaries from Atlas Muni Data, and 3) mortgage-holder race from Home Mortgage Disclosure Act records. These three sources are merged to create a panel of observations at the property-year level. For each home, four pieces of information are observed: (i) the network of taxing entities touching that property, (ii) the annual assessment, (iii) whether any transaction occurs, along with the transacted price if so, and (iv) the race of the homeowner (both buyer and seller in the case of a transaction). For any analysis of assessment ratio, we restrict attention to homes which transact in an arms-length sale with an observed market price, and we focus on the race and ethnicity of the home seller (the homeowner at the time when the assessment was done). We merge this assembled dataset with standard data from the U.S. census and the American Community Survey.

One salient choice we make is to remove all California properties from the final dataset. While taxation in California is legally characterized as an ad valorem tax, the state passed Proposition 13 in 1978, amending the state constitution to place extremely stringent limitations on assessment practices. Assessment growth within each homeowner's tenure is capped at 2% annually, which is far below the growth in market prices in almost every region. In addition, although assessments are supposed to revert to market value upon sale, Proposition 13 also provides a range of mechanisms by which groups of homeowners can transfer the artificially low basis from one property to another. The statewide result has been a decades-long divergence between assessments and market values. As a result, we consider the property tax in California to be ad valorem in name only.[28] Our analysis, applied to the California data, does show similar evidence of racial and ethnic inequality. For completeness, we show this in Section 1.5.1. However, our subsequent analysis of mechanisms in this paper is less relevant for California, simply because assessments there are so mechanically driven by the restrictions of Proposition 13.

1.4.1 Property Records

We obtain property-level records of assessments and transactions from ATTOM. This is a comprehensive dataset with annual observations on 118 million properties in the U.S. from 2003–2016. Assessment and transaction records are sourced from county assessor and recorder offices, respectively. Each property is characterized by a unique identifying ID, which allows us to match assessments with transactions. In addition, each property has a use code which ATTOM harmonizes across local definitions. We restrict our attention to residential properties of up to four units. Commercial property is generally assessed differently from residential properties, so we cannot draw inference from jurisdiction average assessment ratios without restricting to residential properties only. Further, multi-family homes (e.g. large apartment buildings) are sometimes subject to different assessment rules.

[28]Nonetheless, the constitutional amendment authorized by Proposition 13 continues to describe ad valorem property taxation, as in e.g. California Constitution, Article XIII.A Section 1(a).

The restriction to residential properties of one to four units gives us a set of properties which should always be assessed in the same way within jurisdiction. In order to avoid having to impute any market values, our baseline dataset includes only homes for which we observe the sale price in an arm's-length, full consideration transaction. The recorder portion of the ATTOM dataset has several indicator flags for arm's-length transactions and partial interest sales, which collectively can be used to isolate transactions that reflect an accurate signal of market value. The ATTOM data also provides a record of tax dollars paid by the homeowner, along with any exemptions. Importantly, each home is identified with a latitude and longitude for the parcel. These are used to geolocate the home within government borders.

We form assessment ratios using assessments and transactions observed in the same period (year). Assessments applying for year t are produced prior to year t, simply because it takes time for local officials to select, test, and validate their valuation model; and then to deliver notice of assessments to homeowners. Most localities also disseminate revised assessments substantially before tax payments are due to provide for an assessment appeal period (we discuss this at length in Section 1.5.3). Although we are unable to observe the exact date on which the assessment is produced, these practicalities of tax administration give us a high level of confidence that the assessment component of the assessment ratio – the numerator – is not mechanically being affected by the denominator, the realized market price.

1.4.2 Government Boundaries

We obtain shapefiles for government boundaries from Atlas Investment Research's Atlas Muni Data. These 75,000 shapefiles are intended to span the universe of local governments in the U.S. The core set of shapefiles covers counties, cities, towns, schools, and special districts as defined by the U.S. Census. In addition, Atlas Muni Data developed proprietary shapefiles for any entity which has ever accessed public markets, as compiled from Municipal Securities Rulemaking Board filings. Thus, a shapefile is developed for any entity which has ever issued either general obligation or revenue bonds. As debt issuance is very often paired with either broad authority to tax (in the case of general obligation bonds) or a voter-approved one-off tax levy (more common for revenue bonds), we consider each of these entities as a potential taxing entity. Collectively, in addition to the 50 states and D.C., the Atlas data covers 3,142 counties, 46,660 cities or towns, 13,709 independent school districts, and 11,924 special purpose districts.

We use all of these shapefiles to form our taxing jurisdictions. This is a very robust and flexible empirical strategy: if any given entity does not tax, we do not introduce any bias by considering it in forming our unique government networks. If anything, we create another barrier against observing any distortion by restricting our analysis to a (potentially) smaller geographic region. And, of course, if the entity does levy a tax or generate assessments, then failing to take it into consideration would certainly introduce bias.[29] Each shapefile delineates a region in space by connecting a large number of latitude

[29]We suspect it is rare for entities other than counties, cities, or towns to produce assessments; but our

and longitude segments. We use standard GIS techniques to associate each home's longitude and latitude with any shapefile that contains that point. It is an embedded assumption that the latitude and longitude of the property correctly characterizes government association.

We place emphasis on the comprehensive nature of these shapefiles. One significant threat to our research design would be an inability to observe any assessing entity. Practically speaking, because this function so often is assigned to counties or large cities, and because we have shapefiles for every county, and essentially every city and town, we feel that it is very unlikely we have missed such an entity. This is a strength of demonstrating distortions by using assessment ratios. Any statement that we make about tax dollars also has a source of error if we have missed any taxing entity. The breadth of the government shapefiles suggest that any taxing entity not captured in the data is likely to be small.

1.4.3 Home Mortgage Disclosure Act Records

The Home Mortgage Disclosure Act (HMDA) mandates that financial institutions disclose certain information about mortgage applications and mortgage origination at an individual loan level. This law was enacted to provide transparency about credit access for minority residents and within historically redlined neighborhoods. One requirement of the law, therefore, is for financial institutions to solicit and report the racial and ethnic identity of loan applicants. Clients are asked their race and ethnicity directly; the designations are the same as the U.S. Census. A customer can decline to provide this information, and a missing flag is reported as well.[30] HMDA applies to financial institutions meeting certain criteria – the major one being an asset threshold which is currently $46M for depository institutions and $10M for for-profit mortgage lenders. During the 2005–2016 period we consider, between 6,900 and 8,900 institutions reported loans ranging in number from 14.3 to 33.6M annually.[31]

We merge the HMDA records to the ATTOM dataset. This is a fairly standard merge in the literature (see, e.g. Bayer et al. 2017 or Bartlett et al. 2018). HMDA loan records are uniquely identified by: year, census tract, lender name, and dollar amount (rounded to thousands). The ATTOM data contains: transaction date, latitude and longitude of the property, lender name, and dollar amount. We restrict our sample to the highest quality matches, requiring an exact match on year (permitting a one-month overlap between December and January), an exact match on tract, an exact match on (rounded) transaction amount, and a fuzzy string match on lender name.[32]

---

strategy is robust to such an instance.

[30]Regulation C of HMDA also requires loan officers to note race and ethnicity race based on visual observation if the application is made in person and the applicant does not provide the information. During the period covered by this paper, financial institutions were not required to distinguish between application-disclosed information and visually-observed information in reported data.

[31]Summary statistics from www.ffiec.gov.

[32]The diversity of retail-outlet names within a single financial institution can make exact string-matching a challenge in some regions. We rely on a natural language algorithm developed by the Real Estate and Financial Markets Laboratory at the Fisher Center for Real Estate and Urban Economics to match names. The algorithm trains itself within region on perfect singleton matches across all variables other than name, and then uses that mapping to assign a confidence index to each HMDA-ATTOM string-pairing.

The initial merge establishes race/ethnicity of the mortgage-holder at the transaction date.[33] Our end goal is to relate assessment ratios to homeowner race and ethnicity at the time when the assessment was generated. Assessments are produced in advance of the tax year in which they will apply.[34] Therefore, we exploit the dynamic structure of the transactions dataset to build a panel of homes for which we know the declared race and ethnicity of the homeowner at each year. There are two relevant cases: (i) sales and (ii) refinance transactions. For sales, the transaction pins down the race/ethnicity of the buyer, which is then associated with that property in each subsequent year until the next observed transaction. For refinance transactions, we carry race and ethnicity not only forward in time but also backward, as the home does not change ownership. For a large number of transactions, race/ethnicity is not observed.[35] In these cases, we mark race/ethnicity as unknown, and carry that categorization forward and backward in time as appropriate. We fill our panel in this manner, with racial and ethnic indicators updating each time we observe a transaction. As a last step, we remove the observations for which mortgage-holder race and ethnicity is unknown or not declared. We also remove any home which sells in consecutive years. This is because we do not perceive the exact timing of assessment generation. The approach described associates the assessment ratio from a transaction occurring at time t with the race and ethnicity of the homeowner at time $t-1$. Therefore if there are multiple homeowners during year $t-1$, we cannot be sure how to assign race and ethnicity.

Our final baseline dataset is a panel of 6.9M homes. The data are anonymized: each home is characterized by a unique ID variable. For each observation, we have an assessment ratio (comprised of an assessment generated *prior* to sale and the realized sale price observed in a market transaction), know the associated taxing jurisdiction, and have the reported race and ethnicity of the homeowner. Each home is associated with a census tract and a census block group, permitting us to merge in a range of tract-level variables from the American Community Survey five-year estimates.

1.5 Results

The main results of our analysis are organized into five parts.

We first establish the existence and magnitude of the assessment gap in Section 1.5.1. These results document the additional property tax burden faced by an average minority citizen. The analysis is within taxing jurisdiction, which ensures that we are comparing residents who: (i) face the same intended level of taxation, and (ii) are served by the same set of public institutions and government entities. To characterize the distribution of the

[33] HMDA records also include information on co-applicants. We use race and ethnicity of the primary applicant only.

[34] In Philadelphia, for instance, the tax year runs from April 1st to March 31st, with payments due by March 31st. For the 2020 tax year, the Office of Property Assessment mailed notice of assessments to residents at the beginning of April 2019.

[35] This occurs if we cannot match the transaction to a record in HMDA – in the case of a cash transaction, for instance – or if the race/ethnicity is recorded as "not provided" in HMDA.

assessment gap, we present state-level and county-level estimates. We also show that the average assessment gap is increasing in county-level minority population share.

Our second set of results, described in Section 1.5.2, decomposes the assessment gap into two channels. The decomposition is along spatial lines. One channel, which we term "neighborhood composition," relates to spatial variation in the assessment ratio and operates through characteristics of a home's geographic surroundings. We use hedonic regressions to show that market prices and assessed values are well aligned on the valuation of property-level attributes. However, there is large misalignment on pricing of tract-level variables. Even within jurisdiction, people of different races live – on average – in different types of areas. Relative to a black or Hispanic homeowner in the same jurisdiction, the set of local characteristics faced by average white homeowner tend to push market prices up. This residential spatial sorting by race is very well known (Bayer and McMillan, 2005, Logan and Parman, 2017, Lichter et al., 2007, among many others). As would be expected, we show that local attributes are capitalized into market prices. We further show that assessments are much less responsive to these highly-local attributes than market prices are. This suggests that assessors are insufficiently taking neighborhood factors into account when constructing assessments. Attenuated capitalization of these local characteristics therefore leads to under-assessment for the average white resident and over-assessment for the average minority resident.

The other component of the assessment gap is a racial differential that persists even after conditioning away spatial factors. We establish this finding by conducting our analysis within small geographic regions (much smaller than a jurisdiction) to control for between-neighborhood variation. We are implicitly comparing two homeowners of differing race within the same census tract (approximately 4,000 residents) or census block group (approximately 1,200 residents). We refer to this as a "homeowner effect," and posit that a racial difference in assessment appeal outcomes can explain this inequality.

Our third set of results presents evidence on the role of assessment appeals in generating inequality in the property tax. To the best of our knowledge, there is no national compiled dataset on property assessment appeals. In Section 1.5.3, we test the role of assessment appeals in generating inequality, using administrative microdata spanning 12 years of assessment appeals from the second largest county in the U.S. We show racial and ethnic differentials in appeals behavior and outcomes, even within tract and block group. Then we show that assessment patterns nationally are consistent with the appeals channel that we document in a single county.

In Section 1.5.4, we analyze heterogeneity in the assessment gap by racial attitudes and regional minority population share. We first use the measure of racial animus described in Stephens-Davidowitz (2014) to see whether the assessment gap varies with regional racial prejudice. The assessment gap is larger in areas with above-median animus, but is also large and statistically significant in below-median areas as well. This holds both for the overall estimates and the homeowner effect estimates. We also split our sample into quintiles by average county-level minority population share and show that the assessment gap is increasing in minority share.

In Section 1.5.5, our fifth set of results shows that assessment gaps do, in fact, lead to higher tax burdens upon minority residents. While this is the natural implication of assessment ratio distortions – indeed the tight link between assessment ratios and effective tax rates is precisely what motivates our focus on the ratio – we close the loop empirically by demonstrating that this link does hold in the data.

We conclude the Results section with additional discussion of two points: (i) the central role of market prices in our empirical strategy, and (ii) the interplay between income or wealth and the assessment gap.

1.5.1 Assessment Gap Baseline

As outlined in Section 1.2, assessment ratios should be constant within jurisdiction for all residents. As a theoretical statement, this is a necessary condition for an equitable tax benchmark. In practice, the average assessment for any arbitrary grouping of residents should be statistically equal. If the groups in question are distinguished by race (or any other protected class), different group averages represents a discriminatory outcome.

We establish our benchmark finding of an assessment gap by showing that assessment ratios within jurisdiction are, in fact, higher for minority residents. Following equation 1.5, we regress assessment ratio directly on a categorical variable for racial and ethnic groups, along with a jurisdiction-year fixed effect to absorb variation arising from regional scaling choice. Our equitable tax null implies a statistical zero for any race or ethnicity covariate.

Across all our results, we consider three groupings of minority residents. One is mortgage holders whose racial identification in HMDA is "black or African American." The second combines the two largest racial and ethnic minorities in the county: anyone whose racial identification in HMDA is "black or African American" or whose ethnic identification is "Hispanic or Latino."[36] The third is mortgage holders identified in HMDA as having any race other than white or black, and not of Hispanic or Latino ethnicity. This last grouping is not a natural division and masks a large amount of underlying racial heterogeneity. The data is not sufficient to conduct a more precise racial breakdown or a county-of-origin breakdown. We include these results for the sake of completeness. In all cases, the comparison group is non-Hispanic white residents.

Table 1.1 presents our baseline results. Within jurisdiction, assessment ratios are 12.7% higher for black homeowners, 9.8% higher for black or Hispanic homeowners, and just under 3% higher for other nonwhite homeowners. Given a national median effective property tax rate of approximately 1.4%, and a median home value of approximately $207,000, this translates to an additional tax burden of $300–$390 per year for black and Hispanic homeowners.[37]

[36] HMDA regulations do permit loan officers to record race and ethnicity on the basis of visual observation, rather than soliciting the information from the client. Therefore, while this information is often self-reported, that is not always the case.

[37] Averaging over white, non-Hispanic residents, the median jurisdiction in our data realizes an effective tax rate of 1.4%. Other methods of computing a national median property tax rate return similar figures. We obtain a median home value of $207,000 for minority homeowners by taking Zillow's national 2019 estimate of $231,000, and reducing it by 10%, which reflects the ratio of black or Hispanic-owned home value to median home value in our baseline dataset for the latest available year (2016).

We show two results that characterize the distribution of the assessment gap. First, Figure 1.5 shows the assessment gap by state for black residents and for black and Hispanic residents. We present results only from states with at least 500 observations, which excludes seven states. In the remaining set, the assessment gap is positive and strongly statistically significant in most states. For black homeowners, the state level estimates range from 33% to -3%. Estimates are positive and significant in 34 states, positive and insignificant in five, and negative and insignificant in three. For black or Hispanic homeowners, the pattern is very similar.

Second, we estimate the assessment gap at a county level. The results for black residents are shown in Figure 1.6. The distribution for black and Hispanic residents grouped together has a very similar shape.[38] We again restrict attention to counties which have at least 500 observed assessment ratios. This reduces our sample to 671 counties. Our estimates range from 54% to -49%. The interquartile range is 14.8% to 4.7%. Point estimates are positive and significant at the 5% level in 391 counties, positive and insignificant in 219 counties, negative and insignificant in 53 counties, and negative and significant at the 5% level in eight counties. For a black homeowner at the 90th percentile of this distribution, the assessment gap would be 27%. Again considering a $207,000 home subject to a 1.4% tax rate, this would translate into an additional tax burden of $790 annually.

As discussed in Section 1.4, we exclude California from our main analysis. It is widely known that California's property tax has been significantly distorted by Proposition 13, which caps assessment growth at 2% annually during any homeowner's tenure. Under particular circumstances, homeowners can also carry these artificially low assessments from property to property, and can bequeath them to their immediate heirs. In California, for most locations during our sample period, the growth of market prices was significantly greater than 2% annually. For this reason, the primary driver of inequality in the California property tax is more likely related to differentials in homeowner tenure and regional price appreciation, rather than the mechanisms we explore below. For completeness, in Appendix Table A1, we show the results of our baseline analysis applied to California. We do not present other results related to California in this paper.

In Appendix Table A2, we re-estimate the assessment gap using county-year fixed effects rather than jurisdiction-year. The point of this exercise is to show that our careful partitioning of space into taxing jurisdictions is not somehow mechanically driving our results. Differing levels of intended taxation by cities, towns, schools and others makes a within-county analysis of effective tax rate meaningless. However, counties are most often the entity which produces assessments. We can therefore reasonably consider assessment ratio variation within county-year. The results are very consistent with our baseline finding. Inequality in assessment ratios is approximately 4% higher within-county than it is within-jurisdiction. Our preferred specifications all employ the more rigorous within-jurisdiction analysis, not only because it is more likely to hold local assessment practices fixed, but more importantly because jurisdictions are able to hold fixed intended level of taxation and the set of entities providing public services.

Appendix Table A3 shows that our results are unaffected by using jurisdiction-month-

[38]Results are available from the authors upon request.

year fixed effects instead of jurisdiction-year fixed effects. Municipal entities often employ a fiscal year that begins partway through the calendar year (July and October are particularly common starting months). Ideally, our jurisdiction-year fixed effects would align exactly with the fiscal cycle selected by local taxing units, in order to absorb the effect of any deliberate change in assessment practices between fiscal years. We do not observe the local choice of starting month for the fiscal year. Appendix Table A3 shows that our estimates are robust to looking within jurisdiction-month-year. This suggests that any error introduced by forming fixed effects using calendar years rather than the (infeasible) fiscal years does not meaningfully change our estimates.

Finally, we return to the econometric point discussed at the end of Section 1.3. Given large regional heterogeneity in the scaling factor, we use log assessment ratios in all regressions. This does potentially introduce bias in the form of a Jensen's inequality term. In Appendix Table A4, we show that our baseline findings are robust to using simple ratios as the dependent variable. The estimates are naturally lower, as the coefficients are weighted averages of variation around target ratios ranging from 7% to at least 100%. We provide this table only to show that our use of log differences on the left hand side in equation 1.5 is not mechanically driving our results; the estimates themselves have no natural interpretation.

1.5.2 Neighborhood Composition and Homeowner Effect

The preceding section establishes that average assessment ratios differ by race and ethnicity within taxing jurisdiction. Because these regions are carefully constructed to hold taxing units and policy rates fixed, variation in the assessment ratio represents a deviation from an equitable tax benchmark.

Within jurisdiction, there is also a large amount of spatial variation in assessment ratio, and this covaries in striking ways with race. Figures 1–4 show this spatial component in two large counties: Philadelphia County in Pennsylvania, which is coextensive with the city of Philadelphia; and Cook County in Illinois, which contains most of Chicago and several surrounding suburbs. Figure 1.7 is a demographic heatmap of Philadelphia at the census tract level. Using the American Community Survey five-year estimates, we plot the share of black or Hispanic residents in each tract. Figure 1.8 shows within-jurisdiction variation in realized assessment ratios. If the property tax were equitable in Philadelphia, the map in Figure 1.8 would be all the same color. Clearly this is not the case. In addition, there is very high spatial correlation between assessment ratio and minority population share. Figures 1.9 and 1.10 provide a parallel view of Cook County.[39]

These spatial patterns are present to varying extent in many counties and cities. Our decomposition of the assessment gap will disentangle this spatial and geographic (between-neighborhood) variation from the non-spatial (within-neighborhood) drivers. We first establish the magnitude of the assessment gap while holding neighborhood attributes fixed. We show this is 45–50% of the total assessment gap. The remaining 50–55% then is the

[39]Cook County touches numerous towns and other taxing units, and therefore contains multiple jurisdictions. In keeping with our empirical strategy, Figure 1.10 shows demeaned variation within jurisdiction.

between-neighborhood variation. We explore this further in Section 1.5.2.

Homeowner Effect

We proceed by showing that estimates of racial differentials stabilize as we condition on smaller and smaller geography. This exercise approximates the ideal experiment of comparing two homes which are contiguous properties on the same street. Any distortion in assessment ratios arising from neighborhood factors would most plausibly be equivalent for these two homes. Therefore we will describe any remaining difference in assessment ratio as a homeowner effect. We do not observe transactions in sufficient quantity to conduct this analysis using literally adjacent homes. Rather, we will show estimates that appear to converge as we first condition on census tract, and then condition on census block group.

U.S. census tracts are regions of 2,500–8,000 people, with an average of 4,000. Importantly, according to the U.S. Census Geographic Areas Reference Manual, census tracts are initially drawn with the goal of being "as homogeneous as possible with respect to population characteristics, economic status, and living conditions." This criterion provides additional support for our strategy of attempting to hold neighborhood composition fixed by looking within tract. Table 1.2 shows the results of a within-tract analysis.[40] The homeowner effect for black homeowners is 6.4%, for a black or Hispanic homeowner 5.3%, and for other nonwhite homeowners just under 2%.

As noted, the within-tract analysis seeks to absorb variation in spatial characteristics which drive part of the assessment ratio distortion. However, tracts may be large enough that home prices are not identically affected by local factors. The Census further partitions each tract into block groups: regions of 600–3,000 people. This delineation provides an even more defensible setting for our assumption of constant neighborhood characteristics, and so we repeat the preceding exercise at the block group level. We obtain block group shapefiles from the U.S. Census and assign all homes in our sample to their corresponding block group (and jurisdiction in keeping with footnote 40.) Table 1.3 shows the results of a within-block-group analysis. The estimates are fairly stable relative to the tract-level analysis: the point estimates are 5.9% and 4.85% for black and black or Hispanic homeowners respectively; these are both approximately 50bps lower than the estimates in Table 1.2. The point estimate for other nonwhite homeowners is almost the same at 1.9%.

We compare columns (1) and (2) in Table 1.3 with the counterparts in Table 1.1. For black residents, the homeowner effect is 46% of the overall effect. Considering black or Hispanic residents, the homeowner effect is 49% of the total. For the grouping of homeowners who do not identify as white, black or Hispanic, the homeowner effect is 68% of the total. As we describe in the next section, these homeowners on average face a set of neighborhood characteristics most similar to those faced by non-Hispanic homeowners, and accordingly the neighborhood composition effect is small overall.[41]

[40] As always, our analysis is within jurisdiction. Tracts are sometimes split between jurisdictions. Thus, to be precise, we use jurisdiction-tract-year fixed effects.

[41] Again, this grouping obscures a large amount of underlying racial heterogeneity. We include these results for completeness.

Neighborhood Composition

We next explore the portion of the assessment gap which is conditioned away in the preceding analysis by holding spatial factors constant. Figures 1.7–1.10 provide suggestive evidence that racial spatial sorting is relevant for understanding the assessment gap. In each county there is a high tract-level correlation between: (i) highest assessed values relative to market prices and (ii) highest population share of black or Hispanic residents. We establish that this pattern holds in the nationwide data. We estimate the following regression:

$$ar_{icjt} = \beta race_{icjt} + \theta share_{c,jt} + \epsilon_{icjt} \tag{1.8}$$

where ar is the log assessment ratio, i indexes house, j jurisdiction, c census tract, and t year. *Share* is the tract-level population share for a given racial or ethnic group. Fixed effects are again at the jurisdiction-year level. The results are shown in Table 1.4. First, the coefficients on demographic shares are all strongly significant, showing that racial composition correlates strongly with the assessment gap. Second, notice that the direct racial/ethnic coefficients in all columns are much reduced relative to our findings in Table 1.1. In fact, the coefficients are much closer to our estimates of the homeowner effect. This reflects that demographic shares are a strong, though imperfect, statistical proxy for neighborhood factors that correlate with tax-burden variation. We will return to this finding in a number of other ways throughout this section.

A somewhat subtle point is important here. The large estimated effects of demographic shares in Table 1.4 are not, by themselves, evidence of meaningful racial inequality. If white and black residents were evenly spatially distributed, the loading on demographic shares in Table 1.4 would not contribute to average racial disparity.[42] Of course, it is not the case that homeowner location is randomly assigned. In 2017, the average black resident in the U.S. lived in a tract with 43.5% black share, while the average white resident in the U.S. lived in a tract with 7.2% black share.[43] For black or Hispanic residents, the same figures are 56.6% and 17.2% respectively. It is this pattern of residential sorting that, in conjunction with Table 1.4, implies racial and ethnic inequalities linked to spatial factors.

We are not making any causal claim about the estimates in Table 1.4. To the contrary, we will now provide evidence that supports the notion that racial and ethnic shares are a statistical proxy for some latent vector of factors which correlate with the assessment gap. Our baseline findings are group-mean differences in the assessment ratio. Any racially-correlated variable which affects the numerator of this ratio (assessments) differentially from the denominator (market prices) will affect the inequality we measure. Our current aim is to show that many neighborhood-level variables generate variation in the assessment ratio. Then, in the next section, we will use hedonic models to be more precise about which category of variables seems to generate the largest mismatch between the two prices.

From the American Community Survey, we extract a range of variables observable at

[42] If spatial distribution were truly randomly assigned, then by definition any variation in demographic shares would be statistical noise, and the estimated coefficient should be zero. We are making the point that as residential sorting approaches zero, the inequality implied by any loading on demographic shares also approaches zero.

[43] Authors' calculations using American Community Survey data.

the tract-level which would plausibly be assumed to affect house prices. We include these variables in equation 1.8. The results are presented in Table 1.5. To facilitate interpretation, we scale all variables by their standard deviation to show the percent change in assessment ratio correlated with a 1 standard-deviation change in a given variable. Again, the equitable taxation null is that all coefficients should be zero. The coefficients on individual homeowner race, which are very similar to Table 1.4, are shown at the bottom of the regression table.

The surface-level takeaway from this is that many things correlate with the assessment ratio. A positive coefficient represents an increased tax burden, and so Table 1.5 says not only does higher minority population share correlate with higher tax burden, so does lower median income, higher local unemployment rates, and a larger proportion of residents receiving SNAP benefits.[44] Owner percentage and tract level GINI coefficient (a measure of income inequality) are also significantly different from zero. Median age appears to contribute little.

How do we interpret these correlations in the context of the assessment gap? We argue that these patterns arise from market prices being more responsive to neighborhood characteristics than assessed values are. To fix ideas, suppose that assessors impute values as a simple function of home size alone: $A_{icjt} = f(squarefeet_i,\ \#bedrooms_i,\ \#bathrooms_i)$. It is well established in the housing literature that local amenities are also capitalized into home prices (Roback, 1982, Gyourko and Tracy, 1991, Cellini et al., 2010). Thus, if $M_{icjt} = g(squarefeet_i,\ \#bedrooms_i,\ \#bathrooms_i,\ unemployment_{cjt})$, and the market places a nonzero price on local unemployment, then tract-level variation in unemployment will generate variation in the assessment ratio. Further, if the market hedonic price for unemployment is negative, and if unemployment is correlated with minority demographic share (within jurisdiction), then the mismatch will generate an assessment ratio that is increasing in minority share.

The data is consistent with this very simple framework. We establish this by presenting evidence from two hedonic regressions: one with market values as the dependent variable, and the other with assessed valuations as the dependent variable. Specifically, we specify regressions of the form:

$$y_{icjt} = \alpha_{jt} + \beta^y X_{cjt} + \theta W_{icjt} + \epsilon_{icjt} \tag{1.9}$$

where $y \in \{A, M\}$, and i indexes home, j government jurisdiction, c census tract, and t year. X_{cjt} is a vector of tract-level characteristics, and W_{icjt} is a (potentially time-varying) vector of home characteristics including square feet, bedrooms, total rooms, and flags for various amenities. We are interested in comparing the coefficients on β^A with those on β^M. That is, we are interested in knowing whether hedonic characteristics appear to be *differently* capitalized into market valuations and assessed valuations.

Figure 1.11 conveys the results of this analysis. Each bar represents the sensitivity of the (log) assessment ratio with respect to a one standard-deviation change of the given variable. At zero, the assessment hedonic model matches the market hedonics. Above (below) zero, the market hedonic prices are larger (smaller) in magnitude than the corresponding

[44]This is the Supplemental Nutrition Assistance Program, the largest federal nutrition assistance program.

assessment hedonic prices. Finally, bars in black are property-level attributes, and bars in red are tract-level attributes. Figure 1.11 shows that within the context of this hedonic estimation, assessments line up well with market prices on home-level characteristics, but match much less well on neighborhood characteristics. The black bars are all less than 1%: this means that a one standard-deviation shift on any of these dimensions induces less than a 1% shift in the assessment ratio. By contrast, misalignment on tract-level attributes between the assessment and market models is up to an order of magnitude larger. Further, the one variable which receives a greater loading in the assessment model than in the market model is square feet. Table 1.6 shows the estimated hedonic prices from both models. Notice that the signs of the coefficients are all relatively intuitive, with the possible exception of owner share. From columns (2) and (4), we can see that assessors clearly do pay attention to neighborhood characteristics in some manner, but don't place *enough* emphasis thereupon. As a whole, the evidence in Figure 1.11 suggests that assessors: (i) overweight the size of the home, (ii) value other home characteristics fairly precisely, and (iii) underweight local neighborhood composition characteristics.

At a technical level, this underweighting could arise in several ways. All would generate the type of pattern we show here. One possibility is that assessors use hedonic models that include only home attributes and a geographic fixed effect to drive spatial variation in prices. In this case, if the geographic fixed effect is for too broad a region (an entire city, or a quadrant of a city, for example), assessments would be insufficiently high in sub-regions the market values highly and insufficiently low in sub-regions where market prices are low. A similar pattern would result if assessors generate assessments by applying a local growth rate to the prior year's assessment, and that local growth rate is held fixed over a large region (if a single growth rate were picked for an entire city, for example).

While the evidence we provide is consistent with these stories, we are not able to generate a direct empirical test of this hypothesis. Further we cannot disentangle the "mistake" of assessors failing to place enough weight on neighborhood characteristics as the market does, from a deliberate adjustment story where the assessors know how to construct correct valuations but then purposefully distort them in ways that increase the burden on low-income, high-minority, or otherwise economically stressed communities. Empirically speaking, especially at the aggregate level, these would look the same to the econometrician. Either way, the result is a differential in tax burden that arises from neighborhood composition, and creates disparate impact by race.

1.5.3 Mechanism of Homeowner Effect

The neighborhood composition effect is relatively straightforward. In pure economic terms, this looks like evidence that market prices are more efficient in capitalizing amenities and intangibles than nonmarket (administrative) prices. A race-based differential that attaches to individual homeowners is a little more difficult to explain. A natural intuition might be to think of racially biased assessors. We cannot, in fact, rule this out. However, the practical reality of assessments suggests that assessors are unlikely to know the race of the person within any given home. While it is entirely possible that in some smaller regions the property assessor appears at the front door of the home with a clipboard and a checklist, in larger regions there are too many properties to make this practical. Automated Valuation Models or Computer Assisted Mass Appraisal is the standard for larger jurisdictions. The International Association of Assessing Officers seems to be the preeminent professional organization in this space, and publishes professional standard guidelines for mass appraisal (IAOO, 2018). The standards essentially outline multivariate regressions using a relatively small vector of property-level characteristics. It is difficult to think of any reason that an assessor performing mass appraisal of numerous properties would know the race of any given homeowner.[45]

In every jurisdiction of which we are aware, some process for appealing an assessment exists. In general, this tends to be a bureaucratic process run by some agency of local government.[46] A long line of literature in the social sciences suggests a racial component in the extent to which individuals have confidence that public institutions are designed to serve them (extensively surveyed in Nunnally, 2012). This belief may be accurate, or it may be inaccurate but lead to disengagement nonetheless. Therefore, one mechanism we hypothesize and test for is racial differentials in propensity to appeal, likelihood of successful appeals, and degree of reduction conditional on appeal.[47] If one group of residents is more effective at reducing assessment growth by navigating the appeals process, this would lead exactly to the wedge between assessments and transacted values that we observe.

We test this appeals channel for the assessment gap in two ways. First, we use a single large county as a case study. We are unaware of any compiled dataset of appeals at a national level. Although the records are quasi-public, appeals records seem to be posted online less often than the tax rolls themselves. Therefore, we obtain a comprehensive record of appeals submitted to the Cook County Assessors Office between 2002 and 2015, courtesy of Robert Ross (Ross, 2017). Covering 1.9M homes and a population of 5.2M (including the city of Chicago), Cook County is the second largest county in the United States. The Cook County records contain the same anonymized property-ID variable as the ATTOM dataset and therefore is able to be merged directly with our baseline dataset. This yields three additional pieces of information for each property in Cook County: (i) if an appeal was filed in a given tax-year, (ii) whether the appeal was successful, and (iii) if successful, the amount of the reduction.

Cook County has four different channels for appeals: (i) directly through the county assessor's office, (ii) a county board of review, (iii) a state board of review, and (iv) legal

[45] The estimates in Tables 1.2 and 1.3 are already holding neighborhood demographics fixed, so any probabilistic inference about race based on local racial demographics is not a plausible explanation for the homeowner effect.

[46] Our review of state legal codes suggest that two examples are most common: in one case appeals are made directly to a county assessor's office, and in the other case the state empowers some upstream board of review which has authority to adjust the local assessment.

[47] Other scholars have raised this possibility in a property tax setting. Existing work shows a correlation between neighborhood-level demographics and appeal outcomes. Weber and McMillen (2010) also use data from Cook County, covering the period of 2000-2003, along with tract-level demographic data. In a between-tract analysis, they find that high minority share census tracts are correlated with fewer appeal applications and lower success rates. Doerner and Ihlanfeldt (2014) have similar findings in 2005-2009 data from Florida, using a between-block group analysis. To the best of our knowledge, we are the first to use property-level data on homeowner race and ethnicity to conduct a within-neighborhood analysis.

appeal through the Illinois circuit court. Staff at the assessor's office tells us that these latter two are most relevant for commercial properties. The data on residential appeals reflects 3.4M total property-level appeals made through the county assessor's office and through the county board of review. Each record tells us if there is a win at the assessor or board of review level, along with the granted reduction. Staff tell us that usually homeowners appeal first to the assessor's office and then, if unhappy with the assessor's decision, may subsequently pursue the appeal at the county board of review. We are unable to distinguish between a homeowner who accepts a first-stage rejection and one who continues but subsequently loses the appeal at the county board stage. As these two venues are tightly grouped within county administration, we will simply denote a "win" as any homeowner who files an appeal and receives a reduction of any amount, regardless of which office approves the reduction.

While Cook County contains many partially overlapping taxing entities, the county is the only body which produces assessments. We are testing the extent to which appeals can explain the 5-6 percentage points of inequality driven by the homeowner effect, and thus we will conduct our analysis within block-group-year. We are, therefore, comparing appeal propensity, success, and (conditional) magnitude of reduction between two homeowners from the same block group in the same year. Table 1.7 shows the results of this analysis. The estimates in columns (1) and (2) use a linear probability model. The specification in column (3) uses the reduction as a proportion of the proposed assessment as the dependent variable. The baseline rate of appeals in Cook County ranges from 10% to 21% annually during this period, with a mean of 14.6%. The estimate in column (1) shows that black homeowners are 84 basis points less likely to appeal. The baseline success rate for assessment appeals in Cook County ranges from 52% to 80% during this period. The mean is 67.4%. The estimate in column (2) shows that black homeowners are 2.2 percentage points less likely to win, conditional on appealing. The mean reduction granted to a successful appeal in this sample is 12.0%. The estimate in column (3) shows that conditional on a successful appeal, black homeowners receive a reduction smaller by .48 percentage points. Results in Table 1.8 are broadly similar when considering black or Hispanic residents together.

The evidence from this single, large county shows a racial differential in appeals outcomes that will, over time, generate different assessment growth rates. White homeowners will appeal with greater frequency and success, which will generate lower assessment growth relative to black or Hispanic homeowners. Absent other data on appeals, we cannot directly test the assessment appeals channel in other jurisdictions. We can, however, test whether the national data shows evidence of the patterns which this channel would generate. We exploit the time-series structure of assessments in the ATTOM dataset to ascertain whether assessment growth varies by homeowner race or ethnicity. Due to our focus on the assessment ratio, all baseline findings consider homes where we observe an assessment and a market transaction within the same period (year). Assessments are produced annually however, regardless of whether a transaction occurs.[48] Thus we can test for a racial

[48] To the best of our knowledge, property taxes are imposed annually in every jurisdiction. Thus, for purposes of producing a bill, there is an assessment for every tax year; this is what we observe in the

differential in the trajectory of assessments over time.

We will exploit the fact that for a large number of homes in our sample, the racial ownership changes pursuant to a transaction. This permits us to estimate a generalized difference in differences model:

$$y_{icjt} = \alpha_i + \gamma_{cjt} + \beta^r race_{icjt} + \epsilon_{icjt}. \tag{1.10}$$

Each property in this sample is sold at some point. β^r is identified from properties which undergo a change in racial ownership as a consequence of the transaction. Property level fixed effects absorb the between-home variation, and jurisdiction-tract-year fixed effects absorb local housing market variation. The identifying assumption is that within year and within census tract, homebuyer selection into properties is orthogonal to future home price shocks. With this assumption, β is the causal effect of racial ownership on assessments.

Table 1.9 shows the results. As homeowners typically can appeal their assessments each year (or as frequently as new assessments are generated), the channel we posit is most relevant to growth. Accordingly, columns (1) and (2) use the assessment growth (log differences) as the dependent variable. The coefficient in column (1) says that assessment growth is 7bps higher when a black person owns a property, relative to when a white person owns the same property. This is significant only at the 10% level. For black or Hispanic residents the difference in growth is 41bps, and is strongly statistically significant. Given that our sample spans only 13 years, and that an initial transaction is necessary to pin down the race and ethnicity of the homeowner (which further reduces the T-dimension of the usable sample), estimating growth rates may be straining the data even though the sample is very large. In columns (3) and (4), we use (log) levels as the dependent variable instead. The level difference is 29bps and 79bps, respectively. This is consistent with the growth evidence. Within property, assessment levels are higher for minority residents. Given the length of our sample, the estimates in columns (3) and (4) should be thought of as reflecting two to three assessment cycles, which suggests reasonable consistency between the growth estimates and level estimates.

How do these estimates align with the magnitude of the homeowner component of the assessment gap, which is on the order of 5–6%? The more precise (and larger) estimate of column (2) would suggest that the 5% effect (column 2 of Table 1.3) would be generated in 11.5 years. This is a very reasonable figure for median homeowner tenure.[49] The growth differential estimated in column (1), however, would require 80 years to generate the corresponding homeowner effect for black residents. This may suggest that appeals are not the only mechanism in play.

To the extent that assessors do not, in fact, know the race of the homeowner – which is more likely to be the case than not – we argue that Table 1.9 provides strong indirect support for an appeals channel. It is difficult to think of another plausible driver. Any other explanation would require ex-ante racial sorting on *future* assessment growth. The alternate

ATTOM dataset. In jurisdictions that revise assessments less often than annually, the assessment remains static for some number of years (typically 1–2 years, but sometimes longer).

[49]The ACS data implies a median tenure of approximately 12 years.

hypothesis would be in effect that white homebuyers are more likely to select properties that will face a negative assessment shock in the future. We find this less likely than an assessments channel, but acknowledge that the issue remains open for future research.

1.5.4 Heterogeneity by Racial Attitudes, and Demographic Composition, and Homeowner Tenure

It is natural to wonder how the assessment gap relates to racial attitudes. For each channel outlined above, no active expression of bias is necessary, but neither can we rule it out. We use two measures of racial animus developed in Stephens-Davidowitz (2014) to split our sample into regions of high and low racial prejudice. In each sub-sample, we estimate the overall assessment gap and the homeowner effect. The racial animus measures are derived from the regional intensity of Google searches containing the most offensive epithet for African-American people. One measure is produced at the state-level, and the other at the media-market level. For the latter, we use a Nielsen crosswalk to assign the media market measure to counties. We then split our sample along the median of each measure and estimate the assessment gap in a pooled regression. As the measure is designed to capture prejudice towards African-Americans, we estimate the assessment gap only for black homeowners and not for other groupings of minority residents.

Table 1.10 shows the results. Using either measure, the assessment gap is significantly larger in high-animus regions. This holds both in the overall estimates shown in columns (2) and (4), and in the homeowner effect estimates in columns (3) and (5). In regions of below-median prejudice, the assessment gap is still economically and statistically significant. Several mechanisms could lead the assessment gap to be increasing in racial animus. In higher animus regions minority residents may be marginally less likely to appeal property assessments, or less likely to succeed in that appeal. Or, high animus regions may lead to increased racial residential sorting and a larger market-price capitalization of racially correlated factors, which would also lead to the pattern observed here.

We also split the national sample into quintiles based on minority population share at the county-level. The first quintile contains counties with the smallest minority share and the 5th quintile is comprised of counties with the largest. We estimate the assessment gap in each of these sub-samples. Figure 1.12 shows results from these regressions graphically, and Table 1.11 shows the regression estimates. The assessment gap is clearly increasing in minority population share. Since we have shown that a large portion of the assessment gap is linked to spatial sorting, this finding is unsurprising: it has been documented that spatial sorting increases as minority population increases (Card et al., 2008).

Last, we explore the relationship between the assessment gap and homeowner tenure. The evidence on assessment appeals in Section 1.5.3, strongly implies that the assessment gap will increase in homeowner tenure. However, inequality arising through the neighborhood composition channel would not vary with homeowner tenure. Therefore, we would expect a large portion of the assessment gap to remain even while controlling for tenure. This exercise also allows us to rule our the possibility that our results are being driven by any type of unobserved property tax policy which mechanically generates inequality based on tenure - such as California's Proposition 13.[50] The data does not permit us to know the homeowner tenure for our entire sample: for about 40% of the sample, we pin down race and ethnicity using HMDA records from a refinancing transaction, and therefore do not observe original purchase (the transaction data in ATTOM extends back only to 1999). For the remaining 60%, which represents just over 4 million transactions, we observe both the initial purchase and the subsequent sale which generates the assessment ratio. (As noted in Section 1.4.3, assessment ratios are always associated with the race and ethnicity of the home seller, as the seller owns the home at the time when the assessment is generated). This permits us to observe tenure directly. Table 1.12 shows the results of augmenting our baseline assessment gap regression with a control for homeowner tenure. The baseline assessment gap remains large and highly statistically significant. In this subsample of our full data, in fact, the assessment gap is approximately 3 percentage points greater than in the full sample. The assessment gap does increase in homeowner tenure at approximately 50bps per year, which is consistent with the evidence on assessment appeals in Section 1.5.3. Repeated differences in appeals outcomes would lead to an increasing wedge between assessment and market values over time.

1.5.5 Effective Tax Burden

Our last set of results link the assessment gap distortion with actual higher taxation. As a matter of theory, any wedge between assessments and market prices must create a distortion in an ad valorem tax. We are able to observe taxes paid, and therefore can provide the empirical evidence showing that this theoretical relationship does, in fact, hold. Thus far, our focus on assessment ratios has been very deliberate. Assessed values and market prices are observable by the econometrician with little ambiguity. Taxes are more complicated, chiefly due to exemptions.

Every state provides for a variety of tax exemptions in state legislative codes. Most localities have further autonomy to create exemptions. A common example would be a principal residence exemption: Michigan, for example, exempts primary homes from school taxation up to the amount of 18 mills (180bps).[51] Another very common exemption holds for residents of retirement age: New York State permits an exemption of up to 50% for residents over 65 whose income is between \$3,000 and \$29,000.[52] Within these parameters, local units have autonomy to select the precise cutpoints. While these are relatively straightforward, many exemptions are much more complicated. Even at a state level, the list of exemptions tends to be very long and complex. With tens of thousands of local authorities also potentially creating additional exemptions, even observing these exemptions becomes a significant challenge. While the ATTOM data includes a field for exemptions, it is unclear how consistently or accurately this data is reported. We show results: (i) using the reported tax bill directly, and (ii) adding back in the reported exemptions to create a pre-exemption tax bill.

[50] As we discuss in Section 1.4, inequality in California arises directly from homeowner tenure as a byproduct of the stringent assessment caps imposed by Proposition 13.

[51] Michigan Compiled Laws, Section 211.7cc and 380.1211.

[52] https://www.tax.ny.gov/pit/property/exemption/seniorexempt.htm.

Exemptions matter in general because spatial distribution of the exemptions may very well be correlated with racial demographics. If some parts of Florida have more elderly white residents than young black residents, a senior citizen exemption policy would create something that looks like a distortion in the tax burden, but which would be entirely consistent with the legislative intent and public administration of the tax system. We are unable to observe, and thus control for, age of the homeowner – let alone any other individual-level drivers of more complicated exemption policies. The strength of considering the assessment ratio is that none of these confounding factors matter. Using tax dollars paid, we are less able to rigorously strip out potential confounding factors.

Another complicating factor is partial-year tax bills. In some jurisdictions the homeowner of record on a certain date is liable for a full year's worth of property taxes. In others, a partial year of ownership would result in a tax bill spanning only that portion of the year. We do not observe this policy choice at a local level. As we rely upon market prices to determine an effective tax burden, we have another source of bias if race correlates with any propensity to sell in any given year. To provide robustness around this issue, we will compute effective tax burden during the sale year, as well as one year before and one year after sale.

We first estimate the pass-through of the assessment ratio to the effective tax rate. We regress the log effective tax rate on the log assessment ratio. The mechanics of property tax administration would suggest a coefficient of 100%, unless homeowners have not fully exhausted available exemptions. If a region permits homeowners to deduct $5,000 from the assessed value of their primary residence before computing the tax bill and many homes are assessed at less than $5,000 then the pass-through would be less than 100%. Table 1.13 shows these estimated pass-through rates. Column (1) presents estimates for all homeowners in aggregate, and columns (2) and (3) show results by racial and ethnic grouping. Results for black residents alone are very similar, and we do not include them here. Columns (1) and (2) use the actual tax bill. The pass-through is 99%, which closely matches the prediction and suggests that some homeowners are perhaps inframarginal with respect to deductions. Column (3) uses the computed pre-exemption tax bill. Here the estimates are lower. If anything, theory would suggest this number should be closer to 100% than post-exemption figures. We see little reason for the lowered estimates. We take this as an additional reason to be wary of the reported exemption data in ATTOM. Across columns (2) and (3), differences by racial or ethnic identity are not evident.

Table 1.14 directly estimates racial differentials in effective tax rate during the sale year. For black residents, we estimate an effective tax rate that is 14.9% higher in the actual tax bill and 12.2% higher with exemptions added back. This closely brackets the 12.6% excess burden suggested by the assessment gap. Considering black or Hispanic residents together, we find a 11.4% higher effective tax rate from tax bills and 8% increase with observed exemptions added back. Again, this brackets our assessment gap estimate of 9.8%. For the grouping of other nonwhite homeowners, the estimates in columns (5) and (6) again align tightly to our baseline assessment gap estimate. Appendix Tables A5 and A6 show very similar patterns using tax bills one year on either side of the sale.

1.5.6 Additional Discussion: The Role of Market Prices

Market prices are central to our empirical strategy. As discussed in Sections 1.2 and 1.3, this reflects the intention of the property tax as outlined by authorizing legislation. Accordingly, we interpret realized market prices in an arms-length transaction as the appropriate basis for taxation. But what if market prices are "wrong"? A racial differential in transacted prices would also generate a wedge between assessed values and market values, even if assessments perfectly reflect true latent value. It is not immediately obvious how to think about inequality generated by such a mechanism. On the one hand, taxation in general is usually applied to realized financial flows rather than to some latent value. On the other hand, it is very hard to imagine tasking assessors with incorporating homeowner-specific factors that affect market prices. In this section, we provide some evidence on the potential magnitude and direction of inequality driven by racial or ethnic differences in transacted prices.

Several economics papers have explored this possibility. Bayer et al. (2017) uses very similar housing microdata to the ATTOM dataset used in this paper and finds that black and Hispanic *buyers* pay a premium of around 2%. This effect is positive across virtually all racial and ethnic combinations of buyers and sellers, and is largest for within-race transactions (black seller and black buyer; or Hispanic seller and Hispanic buyer). In US housing markets, the majority of transactions occur within-race. Therefore the Bayer et al. (2017) finding would suggest that minority assessment ratios in our sample (which are associated with the race and ethnicity of the home *seller*) may be understated by 2%.[53] In turn, this would imply that racial or ethnic differences in transacted prices *lower* our estimates of inequality by 2%.

Bayer et al. (2017) uses a within-property analysis and restricts attention to four large metropolitan areas to obtain sufficient transaction density. One embedded assumption is that home characteristics stay constant (and are therefore absorbed by the property-level fixed effect). We add additional evidence using a slightly different methodology that relaxes this assumption. While the ATTOM dataset does provide time-varying home characteristics, only macro-level attributes are captured: number of bedrooms, square feet, etc. Assessors typically track major home improvements closely. However, the trigger for updating assessor data is generally a construction permit filed with some local public bureaucracy. We therefore would not observe an indicator for, say, replacing kitchen floor tiles or some other relatively minor improvement which nonetheless would likely impact market price. To address this, we test for racial and ethnic differences in transaction prices which are not predicted by local housing market conditions.

In the set of homes which sell more than once, we define P_0 as the first transaction price. We then form a predicted selling price:

$$\hat{P}_{it} = P_{i0} * \frac{HPI_{zt}}{HPI_{z0}} \tag{1.11}$$

[53] We construct assessment ratios using realized market prices as the denominator. Thus, if realized market prices are higher than "true" value, this would increase the denominator, and reduce the assessment ratio.

where HPI_{zt} is a zip-code level home price index for time t.[54] We then run the following regression:

$$ln(P_{it}) - ln(\hat{P_{it}}) = \gamma_{bg,t} + \beta^r seller\ race_i + \epsilon_{izt} \tag{1.12}$$

where γ_{bg} is a census block group fixed effect. The left hand side is the difference between realized transacted prices and predicted transaction prices. We include a fixed effect at the block group level to account for spatial errors generated by use of a zip-code HPI.[55] The coefficients on our categorical *seller race* variable are estimates of racial and ethnic differences in transacted prices which are not explained by local housing market conditions.

Table 1.15 shows the results. We estimate that black sellers receive 2.2% more than white sellers within the same census block group. Considering black or Hispanic sellers together, the estimated premium is 3.3%. This evidence lines up very closely with the results presented in Bayer et al. (2017). The difference in transacted prices could arise from differential propensity to improve or maintain property, or from a range of housing market frictions. No matter the reason, these results suggest to the extent that a racial differential in market prices exists, realized market prices are slightly higher for minority sellers. This would lead to a *lower* assessment ratio for minority sellers, which means that our estimates of inequality are, if anything, biased downwards on the order of 2–3%.

1.5.7 Additional Discussion: Race vs Income/Wealth

Our baseline estimates of inequality all condition on taxing jurisdiction (annually), but do not include any other control variables. This is intentional. Taxing jurisdictions are formed to create regions where every resident faces the same level of intended taxation. Our equitable taxation null shows that any within-jurisdiction difference in assessment ratios represents an inequality in tax burden. As noted, this is a concept of inequality that aligns tightly to the legal standard of disparate impact. Thus, in this setting, the unconditional difference (within jurisdiction) is the primary statistic of interest.

However, as we have noted, racial and ethnic wealth disparities are among the most persistent and salient stylized facts in household finance. This begs the question of whether the assessment gap simply arises because the U.S. property tax is more regressive than previously understood. This would imply that differences in assessment ratio relate only to income, and that the assessment gap simply appears racially tinged when viewed through the lens of race and ethnicity. We believe the data strongly rejects this notion.

First, our estimates of the homeowner effect are within census block group. This spatial conditioning is a fairly robust non-parametric control for income in many parts of the U.S. Further, our estimates of a racial difference in assessment growth rates are within property. To the extent that choice of home value is also a statistical proxy for income or wealth, this further suggests that the homeowner effect is weakly linked to income.

Our findings on neighborhood composition show that many highly local features are

[54]This use of zip-code level home price indexes holds much in common with the policy approach discussed at greater length in Section 1.6. As in that section, we obtain zip-code HPI measures from Zillow.

[55]As discussed further in Section 1.6, zip codes are relatively large regions.

under-capitalized into assessed valuations, thereby generating inequality in tax burden. Tables 1.5 and 1.6 suggest that neighborhood income is one of these features – at least in a linear specification. So part of our findings do, in fact, suggest that a portion of the inequality we document relates to regressive features of the property tax. However, as Table 1.5 shows, not only are individual racial and ethnic covariates are still large once neighborhood traits are added, minority share is also a highly significant predictor of the assessment gap even after controlling for a range of socioeconomic factors (again, linearly).

We add another piece of suggestive evidence by estimating the assessment gap within ventile of median tract-level income. If the assessment gap were generated primarily by income-related factors, then we should find little evidence thereof within groups equalized by income or wealth. Figure 1.13 shows the estimated assessment gap for black homeowners within income quantile. Two features of this figure are most salient. First, the estimated assessment gap is economically significant within all quantiles. From approximately the median tract through the highest income tract, the level of inequality is fairly stable and is approximately 5%, which matches the magnitude of the homeowner effect. The second salient feature is that the assessment gap is sharply increasing in lower income quantiles. This shows that inequality generated by the assessment gap is heterogenous in income. However, this feature also strongly rebuts the notion that the assessment gap arises primarily from income-related factors. The lowest quantiles in Figure 1.13 are essentially comparing (jurisdictionally demeaned) assessment ratios for the poorest white residents with the poorest black residents. If income were the primary mechanism, these estimates should be close to zero. That they are starkly increasing shows that property assessments are much more misaligned to market values for low-income minority residents than they are for low-income white residents. Figure 1.14 shows the corresponding analysis for black or Hispanic homeowners.

1.6 Policy Corrections

In this section, we discuss a potential approach to address the assessment gap. The inequality we document stems from a wedge between market prices and assessments. Having carefully documented the extent and magnitude of the distortion, it is natural to ask how easily the problem could be fixed. Perhaps it is the case that market prices are so sensitive to geographic variation and property prices so temporally unpredictable, that even the most skilled and attentive assessors office would not be able to equalize tax burdens by racial status. In this section, we show that a relatively simple approach can address a large portion of this inequality.

As more than half of the assessment gap relates to mispricing of local characteristics, we explore whether small-geography home price indexes (HPIs) can be used to reduce inequality. We use zip-code level HPIs to produce imputed assessments, and then compare the racial variation in assessment ratios constructed with our constructed assessments to the variation using true assessments. We find this simple procedure reduces inequality by 55–70%. The average zip code is about twice as large as a census tract. We conjecture that more geographically precise HPIs would be additionally effective in removing assessment ratio variation.

We use publicly available zip-code level HPIs from Zillow to construct assessments. Zillow constructs these HPIs monthly for 15,500 zip codes. This covers 84% of the U.S. population.[56] As some transaction density is needed for a sample size sufficient to produce a reasonable HPI index, these zip codes are highly skewed towards more populous urban areas. The monthly time-series from 1996 can be directly downloaded from Zillow's website at no cost. Zillow began providing these indexes in 2006 and has backwards constructed them to 1996. Zillow has also been increasing its coverage over time.

We construct synthetic assessments using the zip-code HPIs. The algorithm for a synthetic assessment is simple: in any zip code, we take the first observed transaction price and allow this to be the assessment in the month-year of sale. Then we grow that assessment according to the relevant monthly HPI. That is:

$$\hat{A}_{ijzt} = M_{ijz0} \frac{HPI_{ijzt}}{HPI_{ijz0}} \tag{1.13}$$

where 0 denotes the base month-year of the 1st transaction, z denotes zip code, and M_{ijz0} is the observed transaction price in the base year.

We next test the inequality which would be generated by using these synthetic assessments as the basis for property taxation. To do this, we apply the algorithm to carry the synthetic assessment forward in time until we arrive at the month-year of a subsequent transaction. We then form a synthetic assessment ratio at that time t by taking the log difference between our synthetic assessment and the observed transacted price: $\hat{ar}_{ijzt} = log(\hat{A}_{ijzt}) - log(M_{ijzt})$. We evaluate the success of this algorithm for generating assessments by comparing inequality in synthetic assessment ratios to inequality in the realized assessment ratios. Because this simple approach requires two transactions, and is by construction limited to the zip codes that Zillow covers, we end up with a significantly smaller subsample of 2.1M homes. We first document that the assessment gap still exists – and looks similar – in this subsample. Then we document that using synthetic assessments reduces inequality by 55–70%.

The first three columns of Table 1.16 show the assessment gap in the subsample covered by Zillow HPIs. Magnitudes are similar to our baseline findings. The figures in columns (1) and (2) are respectively 1.7% and 1.4% larger than the findings in Table 1.1. For minority homeowners who neither identify as black nor as Hispanic, the estimated effect is nearly the same. Columns (4)–(6) repeat the same regressions using our synthetic assessments. A perfect procedure would produce zeros on the racial and ethnic variables. The synthetic assessments completely reverse the assessment gap, and in fact overshoot. The estimates in columns (4)–(6) of Table 1.16 reflect a *lower* tax burden on minority residents. Of course this is also an inequality in the tax burden. However, the overall distortion is much smaller in magnitude: 4.1% for black homeowners, 5.1% for black or Hispanic homeowners, and effectively zero for other nonwhite homeowners.

Two things are worth emphasizing here. One is that such a straightforward approach is only feasible if some valid HPI exists for small geographic regions. We use Zillow's zip-code

[56] Author's calculations using 2010 decennial census data.

HPIs to demonstrate that inequality can be reduced by using publicly available, easy to obtain data. Zip codes are, however, well known to be formed with little consideration for the institutions and characteristics of the underlying geography. Also, the average zip code contains 9,000 people. This is relatively large: our results suggest that there is meaningful spatial variation between tracts, which are less than half this size on average. We think this is likely to be one important reason that this simple implementation still generates a 4–5% racial difference in assessment ratios. The discussion in Section 1.5.6 also suggests that a racial or ethnic difference in transaction prices could explain 2–3 percentage points of the remaining inequality.

In addition, as a practical matter, assessment values need to be set at the beginning of the tax year, and sales may occur at any time during the next 12 months. Accordingly, racial sorting into areas of higher or lower growth would cause some amount of measured inequality in the realized assessment ratio to arise within the year. To see how important this channel would be, we reproduce a set of synthetic assessments where the assessment is set annually in January of each year. Every transaction then includes up to 12 months of home price growth which is not reflected in the assessment. Appendix Table A7 shows results from this exercise. The estimates are almost unchanged.

The second point of emphasis is that our procedure uses an observed transaction price for the base year value. In order to apply to all properties within a jurisdiction, assessors would need some method for imputing a base-year price for properties which have not sold at any point during the period spanned by the HPI index. Our neighborhood composition findings suggest that this will require assessors to permit prices to vary between small geographic regions. However, racial equity in the initial values is empirically observable and testable. So assessors should be able to iterate a model for initial pricing to land on an equitable distribution of base-year assessments, and then grow those by using some HPI index.[57] The point remains that assessors can make significant strides towards equity by linking assessment growth to small geographic regions within their jurisdiction.

1.7 Conclusion

We document widespread racial inequalities in the U.S. property tax burden. The residential property tax is intended to be an ad valorem levy on the fair market value of the owned asset, yet tax bills are generated as a function of a policy rate and an assessed value. Thus, any wedge between assessed valuations and market prices creates some deviation away from a fair tax benchmark. While local jurisdictions have free choice of a scaling parameter for assessments relative to market prices, the realized ratio within jurisdiction should always be constant across properties. We obtain shapefiles for a comprehensive set of local governments along with other quasi-governmental entities that levy taxes. We associate each home with all governments that contain it. Within these taxing "jurisdictions," defined as a unique set of overlapping governments, the assessment ratio should be constant. Our first

[57] This is, in fact, not particularly dissimilar from the process advocated by IAOO (2018) and other professional guides. However the bulk of this paper serves to show that regardless of process, the outcomes articulated in standards like these are not being widely achieved.

major finding is to document a nationwide assessment gap: assessment ratios are on average higher for minority homeowners. Holding jurisdiction – and thereby public services, intended taxation, and local assessment practices – fixed, the average assessment gap for a black or Hispanic resident in our sample is 10–13%.

We decompose this finding into two components. We show that slightly more than half of the assessment gap can be explained by between-neighborhood variation. Residential sorting by race in the U.S. means that the average black or Hispanic resident faces a different set of local attributes than a white resident does. Market prices appear to be substantially more sensitive to a wide range of observable neighborhood characteristics than assessed valuations. We use hedonic regressions to show that market prices and assessed values align well on home-level attributes, but diverge on tract-level characteristics. This mismatch, along with residential segregation patterns, generates 6–7 percentage points of the total tax burden inequality.

We show that the remaining 5–6 percentage points of inequality persists even within very small geography. We hypothesize that the main channel for this effect is racial differentials in property tax appeals. We use administrative data from Cook County, the second largest county in the US, to demonstrate that such racial differentials can exist: in Cook County, minority residents are 1% less likely to appeal; are 2% less likely to win an appeal; and conditional on success, receive a 2–3% smaller reduction. We then exploit racial changes in ownership around property transactions to test for racial differentials in assessment trajectories, and find patterns consistent with an appeals mechanism in the national data.

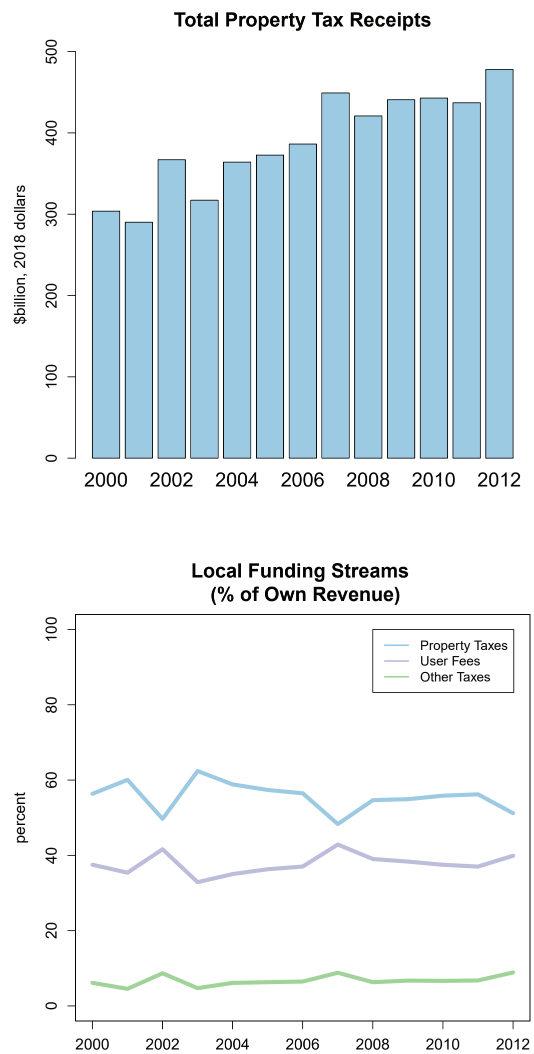
We consider the assessment ratio precisely because distortions in this ratio must mechanically translate to distortions in realized taxation. This is a testable empirical proposition, and we provide evidence to show that this link does in fact hold. Although tax exemptions and sale-related timing complicate any analysis of tax dollars paid, we show that tax bills of minority residents within jurisdiction are higher by roughly the same magnitude as the assessment gap would predict.

Last, we demonstrate that these distortions can be fixed by a relatively simple procedure. Our results suggest that it is important for assessors to recognize that market prices are highly sensitive to local conditions in ways that correlate with race. Accordingly, assessed valuations should reflect price dynamics at a narrow geographical level. We obtain zip-code-level home price indexes. We use these to produce synthetic assessments in the simplest way possible: when a transaction occurs, that becomes the assessment, and from there it evolves in direct proportion with the monthly zip-code HPI. Using a subsample of homes for which we observe two transactions, we create the synthetic assessments at the first sale and model them forward to the second sale. At the second sale, we form a synthetic assessment ratio and use these ratios to test for an assessment gap. The simple synthetic approach reduces the overall inequality by 55–70% and, in fact, flips incidence: the remaining inequality is a lower tax burden on minority residents.

Our baseline findings establish that minority residents in the U.S. face a higher property tax burden than their nonminority neighbors. Although the professional standards for the appraisal industry emphasize that equity in property taxation demands jurisdictionally-constant assessment ratios, the reality of property tax administration in the U.S. is that more jurisdictions fail to achieve this equity than not. Increased taxation clearly represents, in the most literal sense, an incremental cost faced by minority families and an additional impediment to minority wealth building. We know already that there are very striking racial wealth disparities in the U.S. – especially between black and white residents. The inequality we document in taxation is a direct, ongoing, and current source of fiscal headwinds for minority families. We estimate an additional burden of \$300–\$390 per year for the median black or Hispanic family, and up to \$790 for families affected at the 90th percentile of the assessment gap. Residents of local governments implicitly enter a contract agreeing to a given level of taxation in exchange for a bundle of public amenities. Nearly every homeowner in the U.S. faces a property tax, and this large-scale shifting of tax burden onto minority residents violates the notions of equity embedded in the implicit contracts that residents make with local governments.

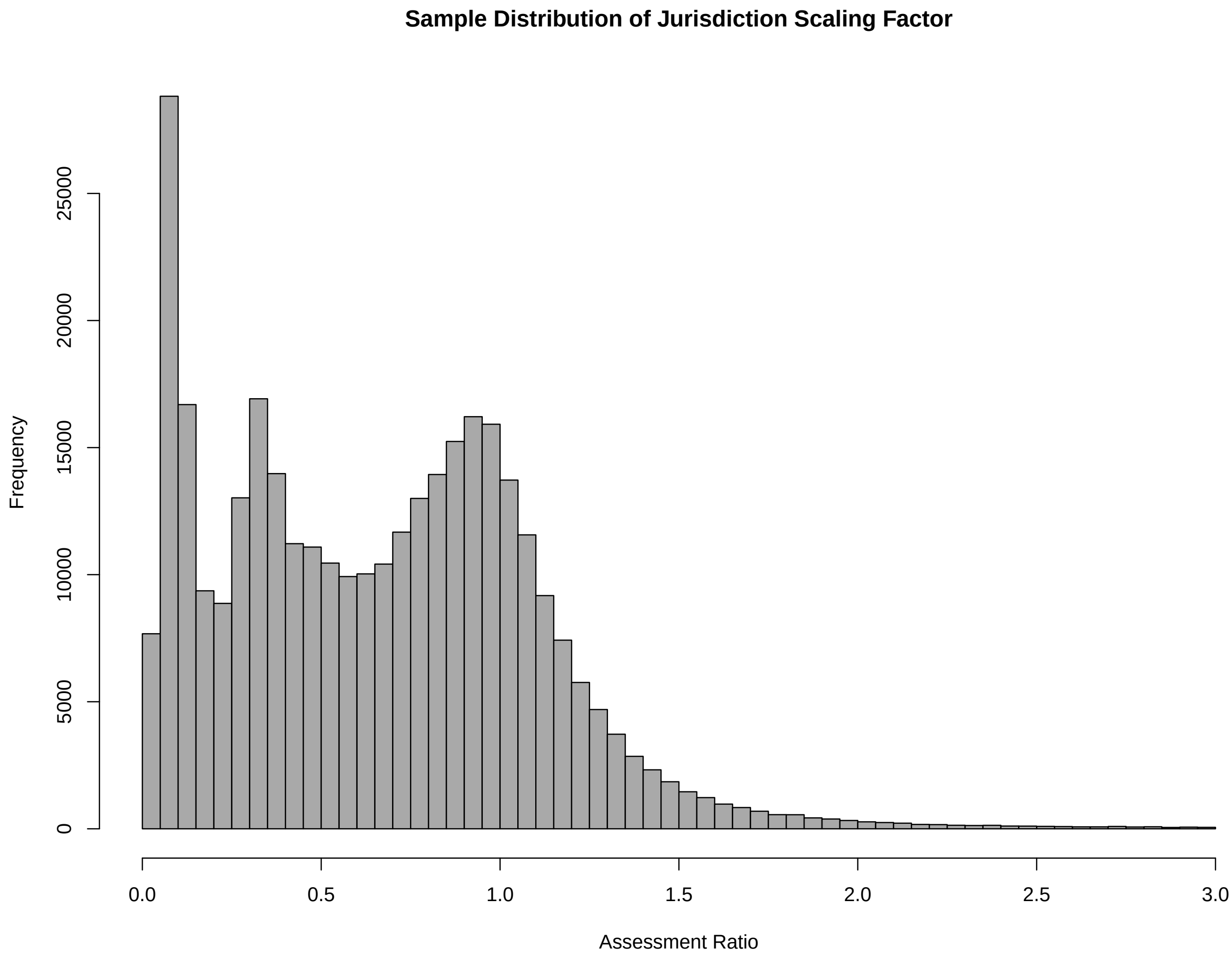
Figures

Figure 1.1: Sources of Funding for Local Governments



Note: This figure shows total property tax receipts (top) for local units of government, and a breakdown of revenue sources (bottom) for 2000 to 2012. Data is from Census of Governments. Full census years are 2002, 2007, and 2012. In all other years, only larger local governments are surveyed. This subset still represents 80-90% of total government budgets, but the omission of smaller governments causes mechanical spikes and dips in 2002, 2007 and 2012 in both figures. The bottom graph shows the average composition of local "own revenue," which is the portion of the budget that the local government can directly affect by policy choices. This excludes intergovernmental transfers from state and federal levels of government. Property tax figures shown are for all property taxes, residential and commercial.

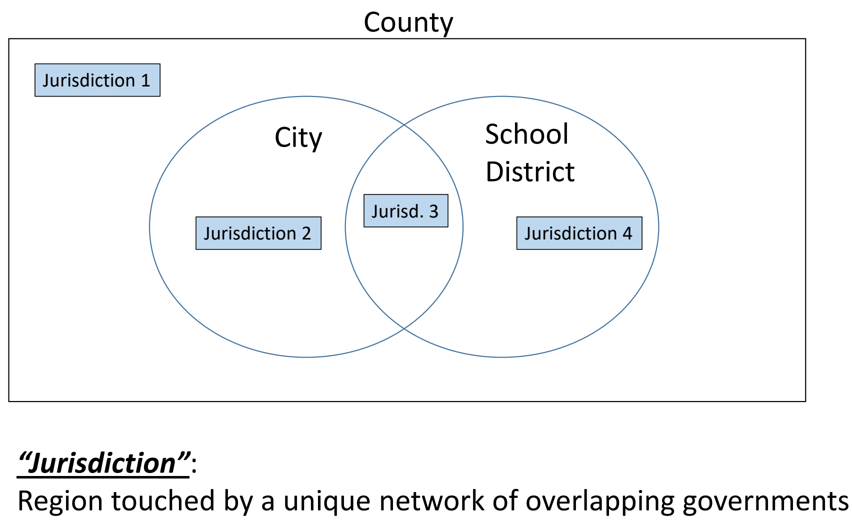
Figure 1.2: Sample Distribution of Local Scaling Factor



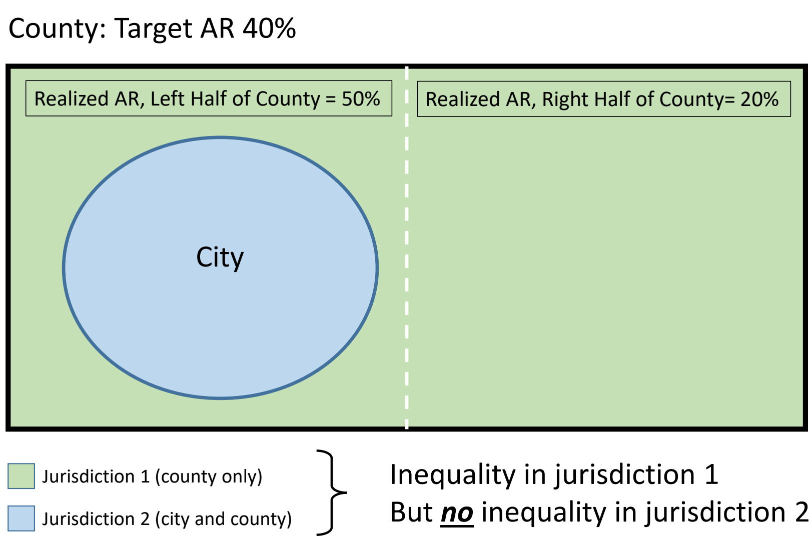
Note: This figure shows the mean *realized* jurisdiction assessment ratio by jurisdiction-year. If assessors make no *average* mistake jurisdiction-wide, then this realized assessment ratio will be equal to the intended scaling factor. In our setting, inequality is a relative difference in assessment ratios between groups within jurisdiction. Therefore, deviations from realized mean are the relevant statistic. If an entire jurisdiction targets a 40% assessment ratio, but realizes a 50% assessment ratio for everyone, this will affect all residents proportionately. Such an outcome may have implications for total revenue raised, but does not represent a source of inequality within jurisdiction. Also note that a jurisdiction-wide wedge between intended and realized scaling factor may not have implications for revenue raised: in many locations the *amount* of intended spending is the politically established choice, and then the aggregate budget is simply divided by aggregate assessed property values to generate a policy tax rate that will be applied to each individual assessment. In this case, if average assessments are higher (lower) across the entire jurisdiction, then implied property tax rates will mechanically be lower (higher) in an exact offset.

Figure 1.3: Taxing Jurisdiction Stylized Examples

Panel A

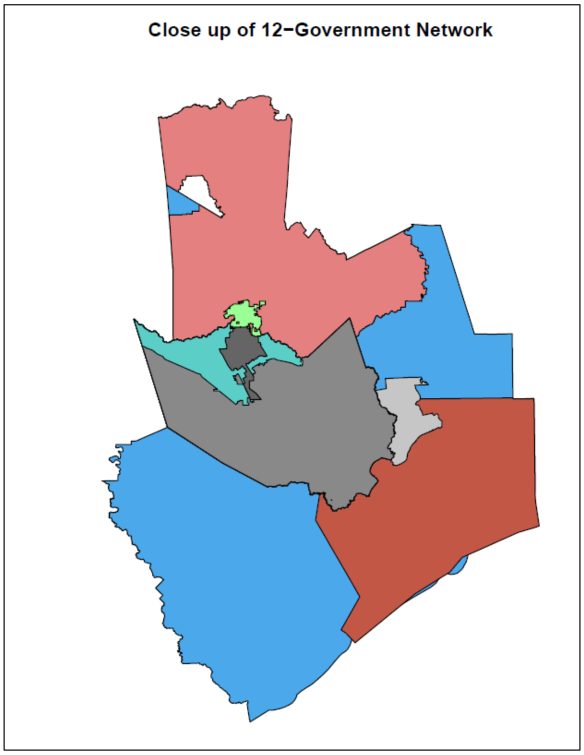


Panel B



Note: This figure shows two examples to illustrate how we form taxing jurisdictions. Panel A shows a stylized example with 3 governments: a county (the large rectangle) which fully contains a city and a school district. The latter two units of government are not spatially coincident. This spatial overlay generates 4 distinct jurisdictions. Panel B presents an example with two governments: the county is again the large rectangle, and a city is entirely contained within the left (blue) portion of the county. In this example, we assume that the county is targeting a 40% assessment ratio, but realizes 50% for every home in the blue region, and realizes 20% for every home in the green region.

Figure 1.4: 12-Government Network in Texas



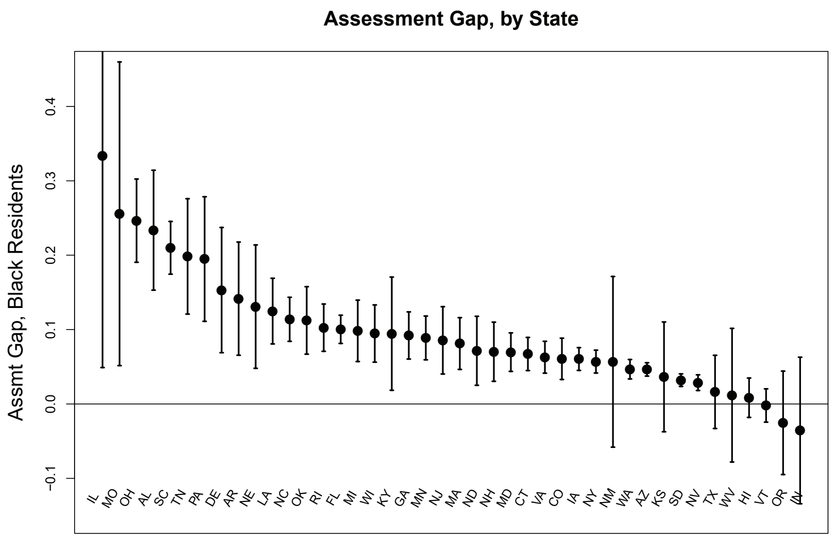
Harris County
City of Houston
Katy Independent School District
Houston Community Colleges
Harris County Flood Control
Port of Houston
Gulf Coast Waste Disposal
Coastal Water Authority
Willow Fork Drainage District
Cinco MUD
North Fort Bend Water Authority
Multi-County Economic Dev. Entity

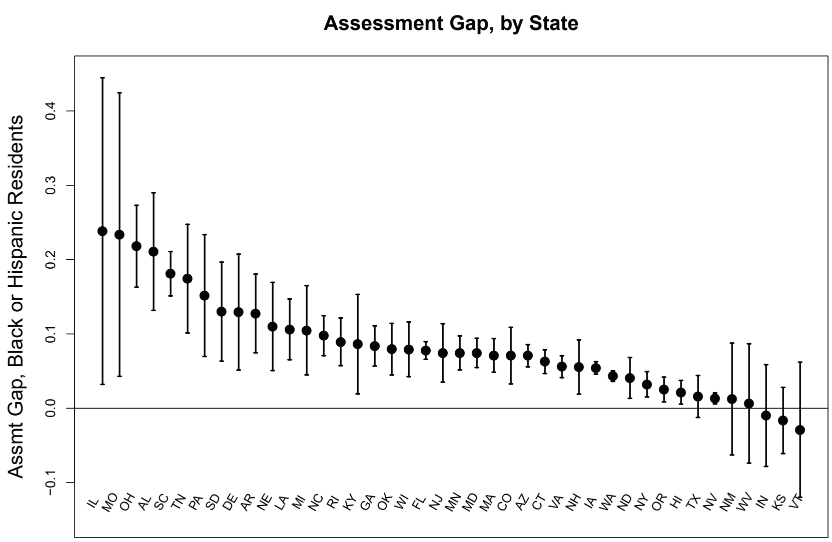
Twelve potential taxing units:

- Every unique combination of these overlapping entities is a jurisdiction
- *One example*: the intersection of all 12 entities

Note: This figure shows the spatial overlay of 12 different local government units in Texas. Some units are proper subsets, and thus fewer than 12 colors are evident in the figure at right. All 12 are listed at upper right. They include "standard" local governments: a county (Harris) and a city (Houston) plus two independent school districts. In addition, there are a range of entities which are related to municipal utilities or economic development initiatives. Each entity listed may, or may not, levy a property tax. Our empirical strategy generates no bias by including an entity as a taxing unit even if it does not, in fact, levy a tax in any particular year. Each unique overlapping combination of these units defines a taxing jurisdiction.

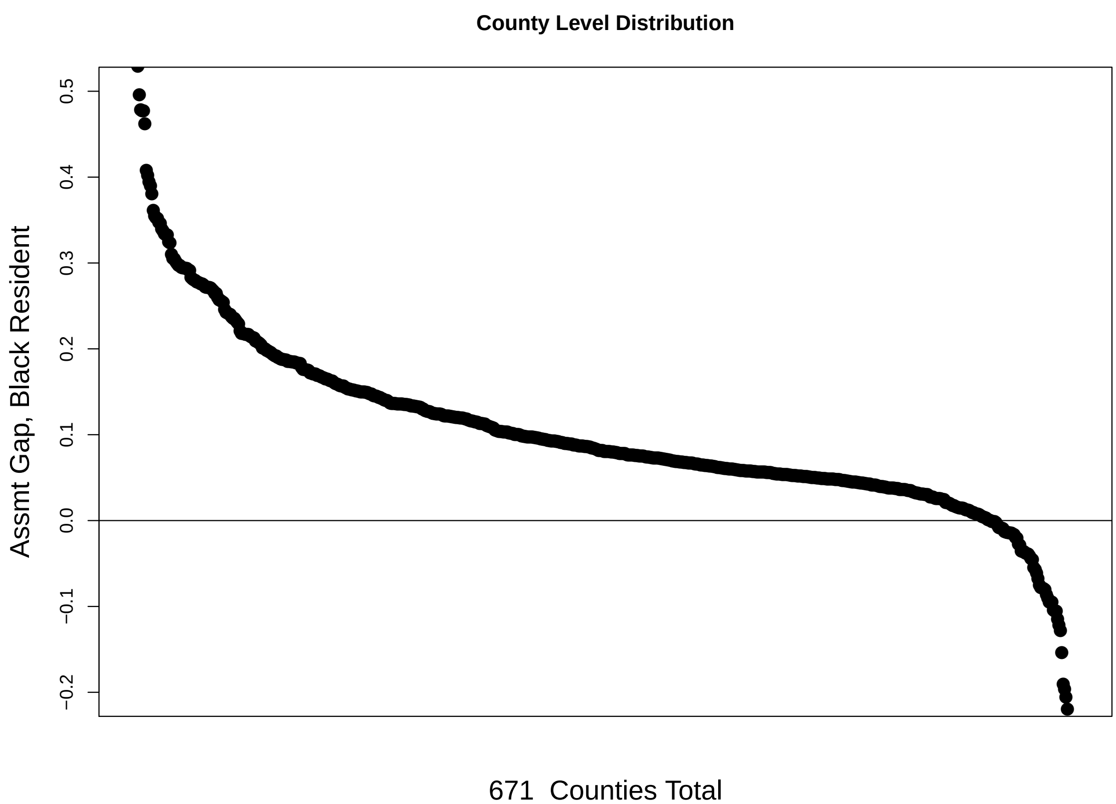
Figure 1.5: State Level Estimates of Assessment Gap





Note: These graphs show state-level estimates of the assessment gap. For every state with at least 500 observations, we regress log assessment ratio on a jurisdiction-year fixed effect and categorical variables for race and ethnicity. The top graph plots the estimated coefficient for black mortgage holders, along with a 95% confidence interval. The reference group is non-Hispanic white residents. Standard errors in the underlying regressions are clustered at the jurisdiction level.

Figure 1.6: County Level Estimates of Assessment Gap



Note: These graphs show county-level estimates of the assessment gap for black residents. For every county with at least 500 observations, we regress log assessment ratio on a jurisdiction-year fixed effect and categorical variables for race and ethnicity. We have sufficient data in 686 counties. We plot the estimated coefficient. For visual clarity, we do not include confidence intervals. Point estimates are positive and significant at 5% in 391 counties, positive and insignificant in 219 counties, negative and insignificant in 53 counties, and negative and significant at 5% in 8 counties. The reference group is non-Hispanic white residents. Standard errors in the underlying regressions are clustered at the jurisdiction level.

Figure 1.7: Philadelphia County/City: Demographic Heatmap

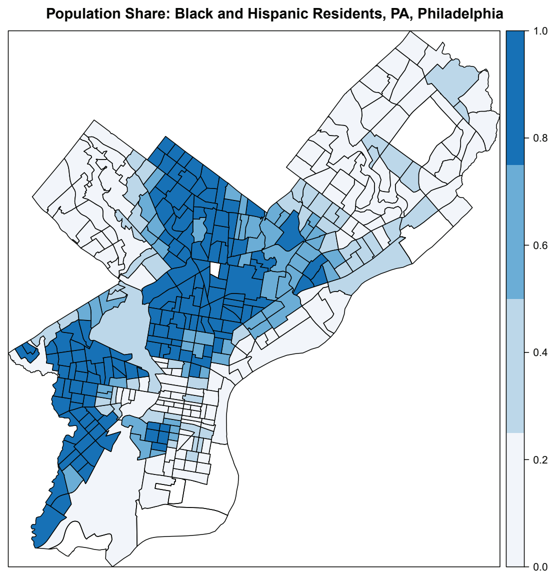
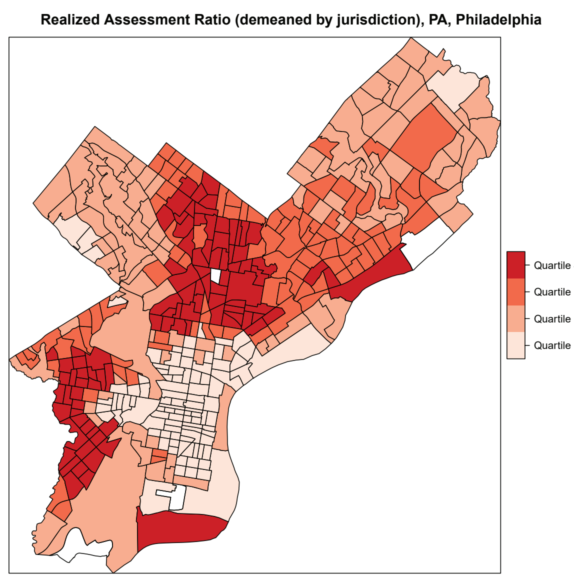


Figure 1.8: Philadelphia County/City: Assessment Ratio Heatmap



Note: Figure 1.7 plots the tract-level share of black and Hispanic residents in Philadelphia, PA using data from the American Community Survey. Tracts having a higher share of black or Hispanic residents are in darker blue. Figure 1.8 shows variation in realized tract-level assessment ratios computed from ATTOM. Realized log assessment ratios are residualized by jurisdiction-year, and then averaged by tract. The result, which we plot in quartiles, is an average proportional deviation from jurisdiction-mean by tract. After absorbing the jurisdiction-year means, an equitable property tax would imply no remaining variation in assessment ratio. Properties within darker red tracts have proportionately greater assessments relative to market price. Because tax bills are computed based on assessments, this mechanically represents a higher tax burden in these areas.

Figure 1.9: Cook County: Demographic Heatmap

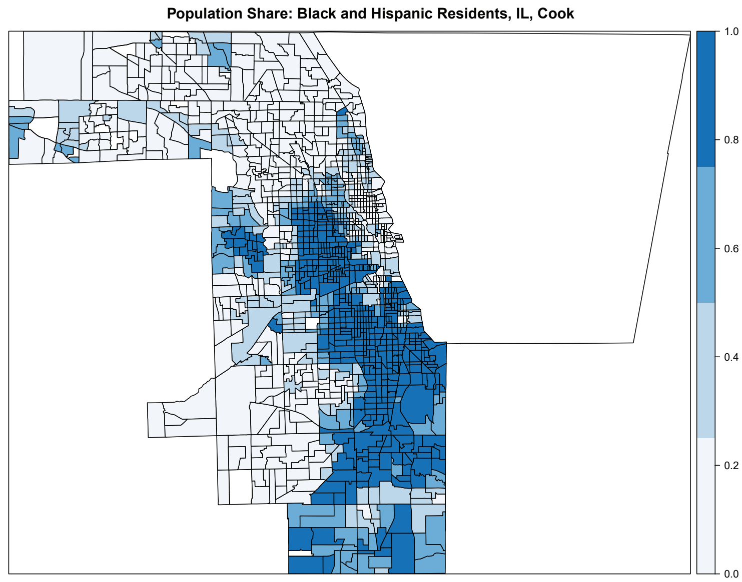
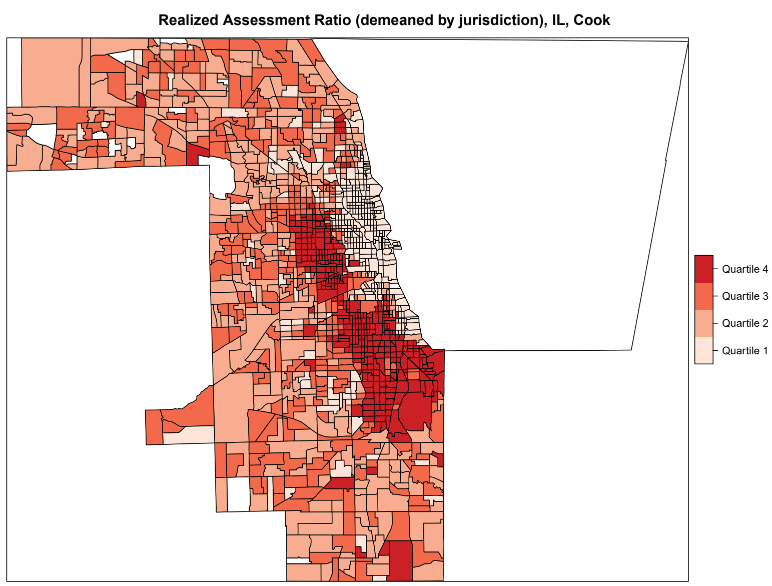
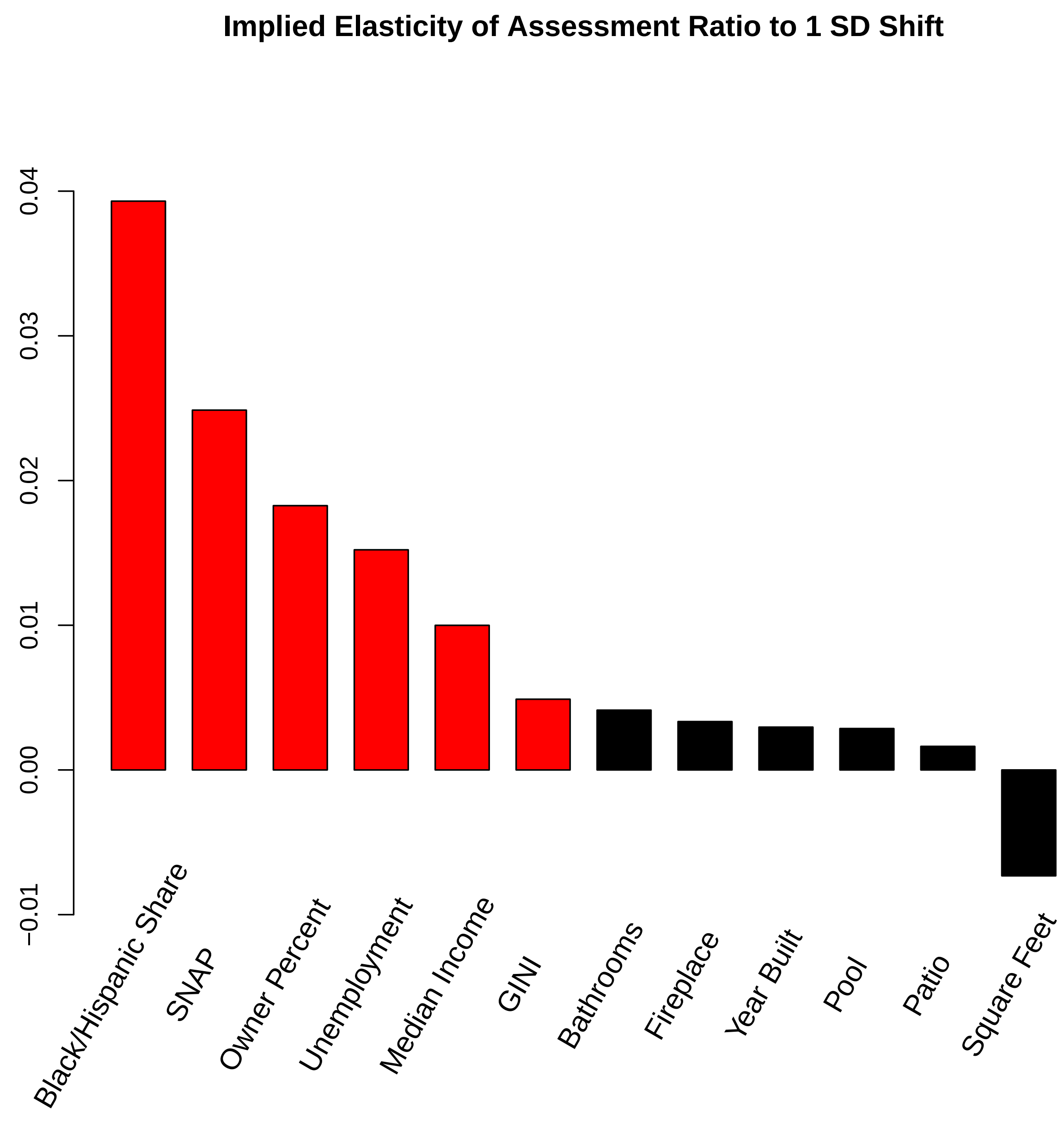


Figure 1.10: Cook County: Assessment Ratio Heatmap



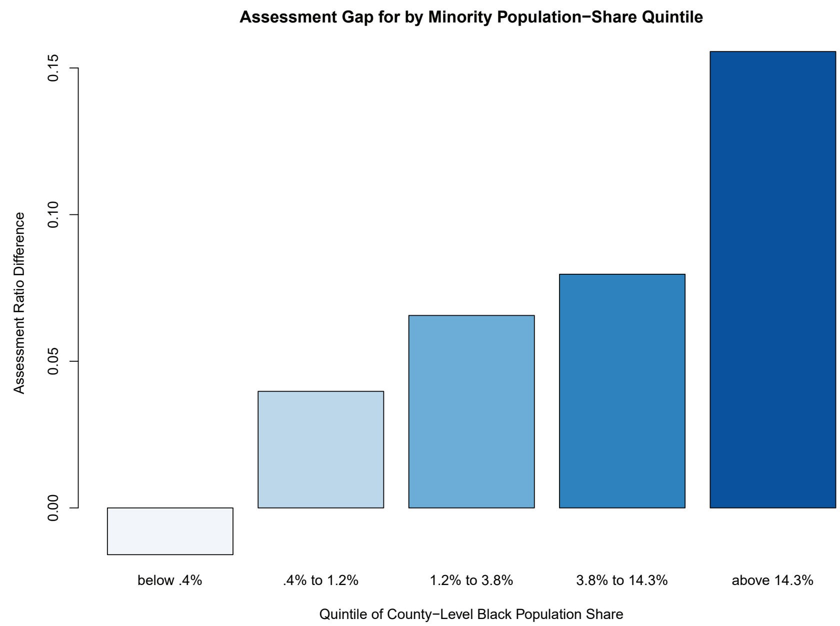
Note: Figure 1.9 plots the tract-level share of black and Hispanic residents in Cook County, IL using data from the American Community Survey. Tracts having a higher share of black or Hispanic residents are in darker blue. Figure 1.10 shows variation in realized tract-level assessment ratios computed from ATTOM. Realized log assessment ratios are residualized by jurisdiction-year, and then averaged by tract. The result, which we plot in quartiles, is an average proportional deviation from jurisdiction-mean by tract. After absorbing the jurisdiction-year means, an equitable property tax would imply no remaining variation in assessment ratio. Properties within darker red tracts have proportionately greater assessments relative to market price. Because tax bills are computed based on assessments, this mechanically represents a higher tax burden in these areas.

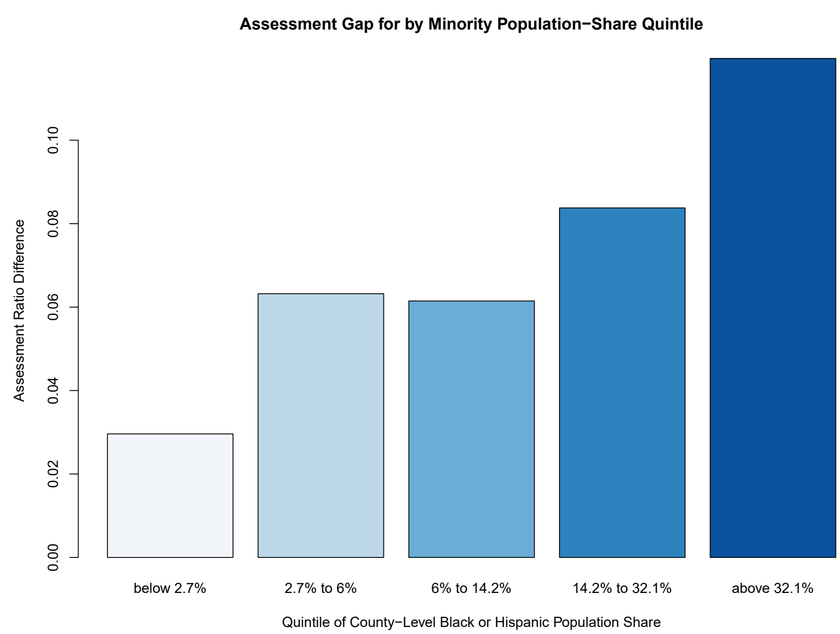
Figure 1.11: Hedonic Models: Mismatch



Note: Each bar in this figure plots the difference between two estimated hedonic prices: one estimated from a model with market values as the dependent variable, and one from a model with assessment values as the dependent variable. Otherwise, the two hedonic models are identical: all regressors are the same. Both market values and assessed values are logged in the underlying models, so the difference between the two estimated hedonic prices represents a proportional shift in the assessment ratio that arises from a one standard-deviation shift in the underlying variable. Bars in red are tract-level characteristics. Bars in black are property-level characteristics. A bar at zero would denote that the market-hedonic is the same as the assessment hedonic price. Larger bars signify a greater disconnect between market-hedonics and assessment-hedonics. Finally, bars above zero denote that estimated *market* hedonic prices are greater in (absolute) magnitude than assessed hedonic prices. Bars below zero denote that the assessment hedonic price is larger. Table 1.6 shows the estimated prices which underlie this figure.

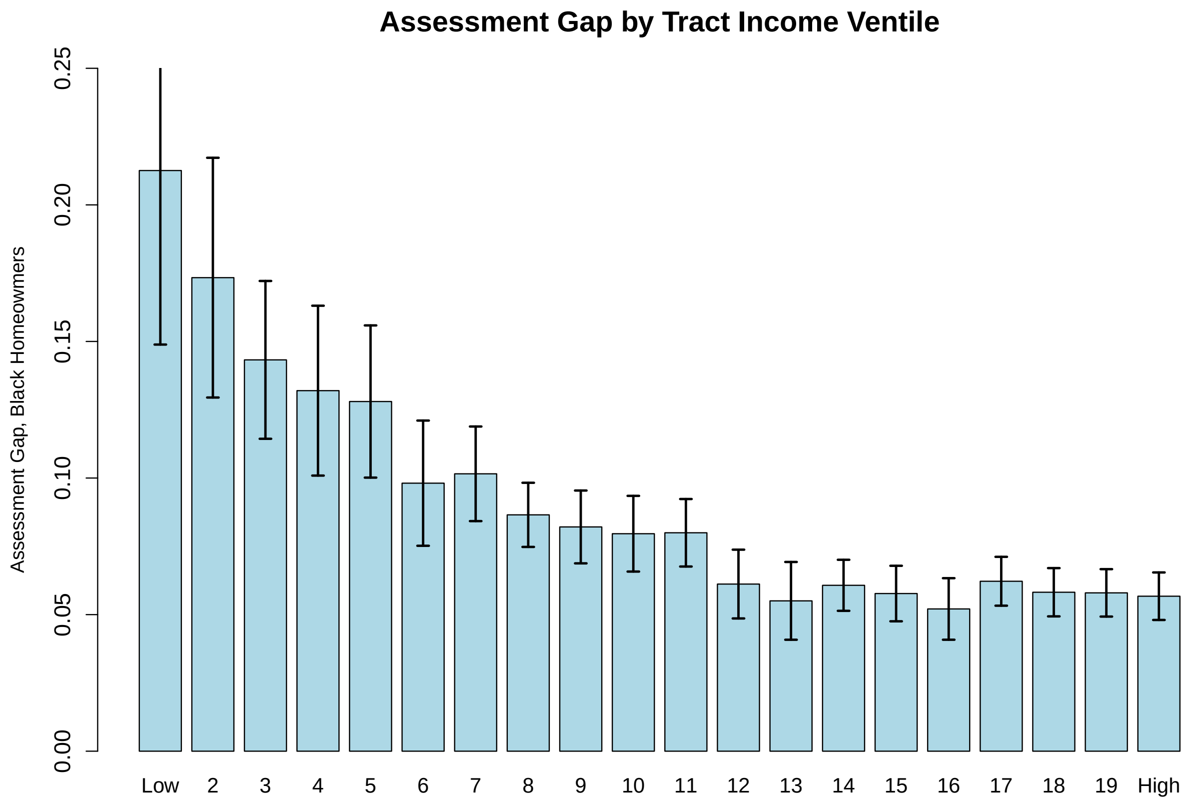
Figure 1.12: Sample Split by County-Level Minority Population Share





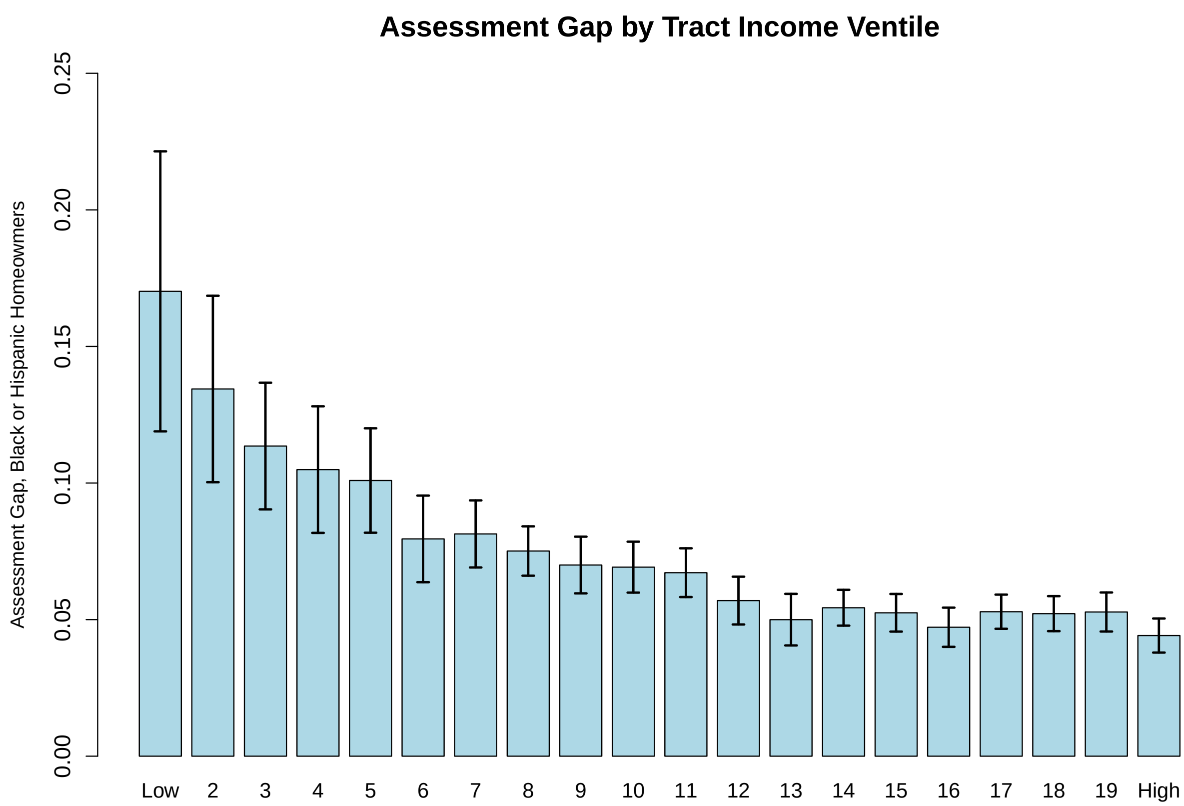
Note: These graphs show results from estimating the assessment gap in sub-samples by minority population share at the county level. We split the sample into quintiles by on average county black or black and Hispanic population share between 2005 to 2016. The quintile range is reflected below each bar. The regression output underlying this table is shown in Table 1.11.

Figure 1.13: Sample Split by Census Tract Median Income



Note: This figure shows the result of estimating the assessment gap for black homeowners in sub-samples by median census tract income. Each property's assessment ratio is residualized on a jurisdiction-year fixed effect. We then graph the mean difference between assessment ratio residuals for black homeowners and white homeowners within each of 20 quantiles based on ACS 5-year estimates of median tract-level income. "Low" denotes tracts having the lowest median income, and "high" denotes tracts with the highest median income. A 95% confidence interval is shown for each income quantile.

Figure 1.14: Sample Split by Census Tract Median Income



Note: This figure shows the result of estimating the assessment gap for black or Hispanic homeowners in sub-samples by median census tract income. Each property's assessment ratio is residualized on a jurisdiction-year fixed effect. We then graph the mean difference between assessment ratio residuals for black/Hispanic homeowners and white homeowners within each of 20 quantiles based on ACS 5-year estimates of median tract-level income. "Low" denotes tracts having the lowest median income, and "high" denotes tracts with the highest median income. A 95% confidence interval is shown for each income quantile.

Tables

Table 1.1: Baseline Assessment Gap

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Mortgage Holder	0.1266***		
	(0.0150)		
Black or Hispanic Mortgage Holder		0.0984***	
		(0.0106)	
Other Nonwhite Mortgage Holder			0.0278***
			(0.0016)
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	37723	37723	37723
Observations	6,987,915	6,987,915	6,987,915
R^2	0.8798	0.8798	0.8798
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: This table shows our baseline findings of a racial assessment gap. We regress the log assessment ratio on a jurisdiction-year fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect an assessment ratio differential for the given grouping of minority residents relative to non-Hispanic white residents. Standard errors are clustered at the jurisdiction level.

Table 1.2: Individual Race Effect: by Tract

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Mortgage Holder	0.0640***		
	(0.0020)		
Black or Hispanic Mortgage Holder		0.0530***	
		(0.0015)	
Other Nonwhite Mortgage Holder			0.0198***
			(0.0006)
Fixed Effects	Jurisd-Tract-Yr	Jurisd-Tract-Yr	Jurisd-Tract-Yr
No. Clusters	37723	37723	37723
Observations	6,987,915	6,987,915	6,987,915
R^2	0.9005	0.9005	0.9005
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the within-tract portion of the assessment gap. We regress the log assessment ratio on a jurisdiction-tract-year fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect an assessment ratio differential for the given grouping of minority residents relative to non-Hispanic white residents. Standard errors are clustered at the jurisdiction level.

Table 1.3: Individual Race Effect: by Block Group

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Mortgage Holder	0.0588***		
	(0.0019)		
Black or Hispanic Mortgage Holder		0.0485***	
		(0.0014)	
Other Nonwhite Mortgage Holder			0.0190***
			(0.0007)
Fixed Effects	Jurisd-BG-Yr	Jurisd-BG-Yr	Jurisd-BG-Yr
No. Clusters	37723	37723	37723
Observations	6,987,915	6,987,915	6,987,915
R^2	0.9166	0.9166	0.9166
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the within-block group portion of the assessment gap. We regress the log assessment ratio on a jurisdiction-year-block group fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect an assessment ratio differential for the given grouping of minority residents relative to non-Hispanic white residents. Standard errors are clustered at the jurisdiction level.

Table 1.4: Race and Demographic Shares

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Mortgage Holder	0.079***		
	(0.004)		
Black Share	0.299***		
	(0.046)		
Black or Hispanic Mortgage Holder		0.067***	
		(0.003)	
Black or Hispanic Share		0.277***	
		(0.042)	
Other Nonwhite Mortgage Holder			0.029***
			(0.002)
Other Nonwhite Share			−0.139*
			(0.083)
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	37679	37679	37679
Observations	6,944,439	6,944,439	6,944,439
R^2	0.881	0.881	0.880
Note:		*p<0.1; **p<0.05; ***p<0.01	

Note: This table augments our baseline assessment gap findings in Table 1.1 with one measure of spatial variation: tract-level demographic shares. We regress the log assessment ratio on a jurisdiction-year fixed effect, categorical groupings by racial and ethnic identity, and tract-level demographic shares from the American Community Survey. In all columns, the reference group for mortgage holder race and ethnicity is non-Hispanic white residents, and for clarity other mortgage holder coefficients are not reported. The mortgage holder coefficients in this table reflect an assessment ratio differential for the given grouping of minority residents relative to non-Hispanic white residents. The share coefficients represent additional variation in the assessment ratio that correlates with demographic composition of the surrounding tract, holding mortgage holder race fixed. Standard errors are clustered at the jurisdiction level.

Table 1.5: All Neighborhood Correlates

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Share	0.027***		
	(0.005)		
Black or Hispanic Share		0.035***	
		(0.006)	
Other Nonwhite Share			−0.005
			(0.004)
Median HH Income	−0.021***	−0.015***	−0.024***
	(0.005)	(0.004)	(0.004)
Unemployment	0.015***	0.017***	0.020***
	(0.004)	(0.004)	(0.005)
SNAP Assistance	0.033***	0.030***	0.040***
	(0.004)	(0.003)	(0.005)
Owner Percentage	0.021***	0.020***	0.022***
	(0.004)	(0.004)	(0.004)
GINI Coef	−0.011***	−0.009***	−0.012***
	(0.002)	(0.002)	(0.002)
Median Age	0.003*	0.008***	0.002
	(0.002)	(0.003)	(0.002)
Homeowner Race Coef	0.077	0.065	0.023
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	37679	37679	37679
Observations	6,944,439	6,944,439	6,944,439
R^2	0.881	0.881	0.881
Note:	*p<0.1; **p<0.05; ***p<0.01		

Note: This table augments our baseline assessment gap findings in Table 1.1 with several measures of spatial characteristics. All regressors are tract-level variables from the American Community Survey 5-year estimates. Standard errors are clustered at the jurisdiction level. We continue to hold homeowner race fixed in this regression: those coefficients are reported in the first line of notes immediately under the estimated coefficients. Standard errors are clustered at the jurisdiction level.

Table 1.6: Hedonic Prices

	Market	Assessment	Market	Assessment
	(1)	(2)	(3)	(4)
Black Share	-0.092^{***}	-0.056^{***}		
	(0.004)	(0.004)		
Black or Hispanic Share			-0.117^{***}	-0.078^{***}
			(0.006)	(0.005)
Median HH Income	0.157^{***}	0.144^{***}	0.145^{***}	0.135^{***}
	(0.008)	(0.008)	(0.008)	(0.008)
Unemployment	-0.027^{***}	-0.013^{***}	-0.030^{***}	-0.015^{***}
	(0.003)	(0.002)	(0.004)	(0.002)
SNAP Share	-0.089^{***}	-0.061^{***}	-0.075^{***}	-0.050^{***}
	(0.006)	(0.004)	(0.006)	(0.004)
Owner Share	-0.049^{***}	-0.032^{***}	-0.053^{***}	-0.035^{***}
	(0.005)	(0.003)	(0.005)	(0.004)
GINI	0.066^{***}	0.059^{***}	0.058^{***}	0.053^{***}
	(0.004)	(0.004)	(0.004)	(0.004)
Square Feet	0.256^{***}	0.264^{***}	0.256^{***}	0.264^{***}
	(0.029)	(0.030)	(0.029)	(0.030)
Bathrooms	0.107^{***}	0.103^{***}	0.107^{***}	0.103^{***}
	(0.017)	(0.017)	(0.017)	(0.017)
Year Built	0.031^{***}	0.028^{***}	0.030^{***}	0.028^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Other Attributes	Y	Y	Y	Y
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	26152	26152	26152	26152
Observations	4,877,658	4,877,658	4,877,658	4,877,658
R^2	0.773	0.942	0.773	0.942

Note: *p<0.1; **p<0.05; ***p<0.01
Not shown: coefficients on indicators for patio, pool, and fireplace

Note: This table reports estimated hedonic prices from two separate hedonic models. The first model uses (log) market as the dependent variable. These estimates are reported in columns 1 and 3. The second model uses (log) assessed values as the dependent variable. These estimates are reported in columns 2 and 4. Otherwise, the two hedonic models are identical: all regressors are the same. The table omits estimated coefficients for indicator variables stating whether a property has a patio, pool, or fireplace. Standard errors are clustered at the jurisdiction level. Figure 1.11 shows the difference between attribute-coefficients graphically.

Table 1.7: Cook County Appeals

	Dependent Variable:		
	Appeal	Win Appeal	Reduction
	(1)	(2)	(3)
Black Mortgage Holder	-0.840^{***}	-2.193^{***}	-0.480^{***}
	(0.083)	(0.354)	(0.117)
Baseline Rate	14.6	67.4	12.0
Fixed Effects	BG-Year	BG-Year	BG-Year
No. Clusters	3954	3933	3893
Observations	4,076,655	694,553	476,368
R^2	0.383	0.415	0.442
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table uses administrative microdata on property tax appeals in Cook County. The first column shows unconditional propensity to appeal. Column 2 conditions on a homeowner having filed an assessment appeal. Column 3 conditions on a successful appeal. In columns 1 and 2, the dependent variable is a binary indicator. In column 3, the dependent variable is the reduction amount divided by the proposed assessment. Fixed effects across all columns are at the block-group-year level. Standard errors are clustered at the block group level.The baseline rates for (i) appeal propensity, (ii) winning appeal, and (iii) reduction conditional on a successful appeal are reported in the first line below the estimates. Coefficients and baseline rates are reported as percents.

Table 1.8: Cook County Appeals

	Dependent Variable:		
	Appeal	Win Appeal	Reduction
	(1)	(2)	(3)
Black or Hispanic Mortgage Holder	-0.982^{***}	-1.993^{***}	-0.258^{***}
	(0.068)	(0.245)	(0.074)
Baseline Rate	14.6	67.4	12.0
Fixed Effects	BG-Year	BG-Year	BG-Year
No. Clusters	3954	3933	3893
Observations	4,076,655	694,553	476,368
R^2	0.383	0.415	0.443
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: This table uses administrative microdata on property tax appeals in Cook County. The first column shows unconditional propensity to appeal. Column 2 conditions on a homeowner having filed an assessment appeal. Column 3 conditions on a successful appeal. In columns 1 and 2, the dependent variable is a binary indicator. In column 3, the dependent variable is the reduction amount divided by the proposed assessment. Fixed effects across all columns are at the block-group-year level. Standard errors are clustered at the block group level. The baseline rates for (i) appeal propensity, (ii) winning appeal, and (iii) reduction conditional on a successful appeal are reported in the first line below the estimates. Coefficients and baseline rates are reported as percents.

Table 1.9: Effect of Racial Ownership on Assessments

	Assessments			
	Growth		Levels	
	(1)	(2)	(3)	(4)
Black Mortgage Holder	0.0711*		0.2917***	
	(0.0386)		(0.0415)	
Black or Hispanic Mortgage Holder		0.4103***		0.7923***
		(0.0255)		(0.0274)
Fixed Effects	Two-Way	Two-Way	Two-Way	Two-Way
No. Clusters	12268641	12268641	12268641	12268641
Observations	54,970,191	54,970,191	54,970,191	54,970,191
R^2	0.6925	0.6925	0.9910	0.9910
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results of a generalized difference-in-differences estimation. The dependent variable is logged assessment value. Every home in this sample is transacted at least once. Fixed effects are two-way: property and tract-year. In columns 1 and 2, the dependent variable is growth rates (log difference in assessed value). In columns 3 and 4, the dependent variable is the logged assessment. Standard errors are clustered at the property level.

Table 1.10: Sample Split by Racial Attitudes

	Assessment Value / Market Value				
	Baseline	By Media Market		By State	
	(1)	(2)	(3)	(4)	(5)
Black Mortgage Holder	0.128***				
	(0.015)				
Black, High Animus		0.150***	0.070***	0.145***	0.076***
		(0.022)	(0.003)	(0.011)	(0.003)
Black, Low Animus		0.084***	0.055***	0.106***	0.049***
		(0.008)	(0.002)	(0.033)	(0.002)
Fixed Effects	Jurisd-Yr	Jursid-Yr	Jurisd-Tract-Yr	Jurisd-Yr	Jursid-Tract-Yr
No. Clusters	37106	37106	37106	37106	37106
Observations	6,856,585	6,856,585	6,856,585	6,856,585	6,856,585
R^2	0.881	0.881	0.902	0.881	0.902
Note:					*p<0.1; **p<0.05; ***p<0.01

Note: This table shows results of using the measures of racial animus described in Stephens-Davidowitz (2014) to split our sample into regions of above- and below-median prejudice. Column 1 shows baseline results before splitting the sample. Columns 2 and 3 use a media-market measure of animus. We use a Nielsen crosswalk to associate media markets with individual counties. Columns 4 and 5 use a state-level measure of animus. For each measure, the first result (column 2 or 4) shows the overall assessment gap. The second result shows the homeowner effect estimated within jurisdiction-tract-year. For all specifications, standard errors are clustered at the jurisdiction level.

Table 1.11: Sample Split by County-Level Minority Population Share

Panel A

	Assessment Value / Market Value Quintile of County-Level Minority Population Share				
	(1)	(2)	(3)	(4)	(5)
Black Mortgage Holder	−0.016	0.040***	0.066***	0.080***	0.156***
	(0.054)	(0.007)	(0.004)	(0.006)	(0.022)
Fixed Effects	Jurisd-Yr	Jurisd-Yr	Jurisd-Yr	Jurisd-Yr	Jurisd-Yr
No. Clusters	2008	6491	9490	12813	6323
Observations	53,919	405,323	909,640	3,114,742	2,372,961
R^2	0.856	0.938	0.906	0.888	0.850
Note:					*p<0.1; **p<0.05; ***p<0.01

Panel B

	Assessment Value / Market Value Quintile of County-Level Minority Population Share				
	(1)	(2)	(3)	(4)	(5)
Black or Hispanic	0.030**	0.063***	0.061***	0.084***	0.120***
Mortgage Holder	(0.014)	(0.006)	(0.003)	(0.006)	(0.019)
Fixed Effects	Jurisd-Yr	Jurisd-Yr	Jurisd-Yr	Jurisd-Yr	Jurisd-Yr
No. Clusters	3215	5989	10998	12089	4843
Observations	73,243	295,057	1,433,767	2,796,141	2,258,377
R^2	0.819	0.786	0.858	0.879	0.882
Note:					*p<0.1; **p<0.05; ***p<0.01

Note: Each panel shows the results from estimating the assessment gap on sub-samples based on county-level demographics. For Panel A, we split our baseline sample into quintiles by average county black population share. In Panel B the sample is split by black or Hispanic population share. In each panel, column 1 shows the estimated assessment gap within the lowest minority-population quintile, and column 5 shows results for the highest quintile. Regressions are run separately rather than pooled. We include jurisdiction-year fixed effects in all specifications. Standard errors are clustered at the jurisdiction level.

Table 1.12: Assessment Gap with Homeowner Tenure

	log(Assessment) - log(Market)	
	(1)	(2)
Black Mortgage Holder	0.1532***	
	(0.0183)	
Black or Hispanic Mortgage Holder		0.1175***
		(0.0122)
Years Since Purchase	0.0050***	0.0053***
	(0.0003)	(0.0003)
Fixed Effects	Jurisd-Year	Jurisd-Year
No. Clusters	32567	32567
Observations	4,117,014	4,117,014
R^2	0.8937	0.8937
Note:	*p<0.1; **p<0.05; ***p<0.01	

Note: This table augments our baseline assessment gap analysis with the additional linear covariate of homeowner tenure. We regress the log assessment ratio on a jurisdiction-year fixed effect, categorical groupings by racial and ethnic identity, and the years since purchase. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. Standard errors are clustered at the jurisdiction level.

Table 1.13: Assessment Ratio Pass Through to Tax Bill

	Effective Tax Rate - Year of Sale (%)		
	Tax Bill	Tax Bill	Before Exemptions
	(1)	(2)	(3)
All Mortgage Holders	0.9913***		
	(0.0039)		
White Mortgage Holder		0.9925***	0.8569***
		(0.0037)	(0.0128)
Black or Hispanic Mortgage Holder		0.9857***	0.8517***
		(0.0056)	(0.0131)
Other Nonwhite Mortgage Holder		0.9892***	0.8536***
		(0.0040)	(0.0131)
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	26371	26371	26371
Observations	3,373,164	3,373,164	3,373,164
R^2	0.9191	0.9192	0.7672
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results of regressing log effective tax rate on log assessment ratio. Column 1 presents estimates for all homeowners. Columns 2 and 3 show a breakdown by racial and ethnic grouping. Results for black homeowners alone are very similar to those reported here. In columns 1 and 2, the dependent variable is an effective rate formed using the actual tax bill reported in the ATTOM dataset. Column 3 computes a pre-exemption effective rate by adding reported exemptions back to the reported tax bill. The effective rate is computed by using the tax bill reported in the same year as the sale. All specifications use jurisdiction-year fixed effects. Standard errors are clustered at the jurisdiction level.

Table 1.14: Effective Tax Rate, Sale Year

	Effective Tax Rate - In Sale Year (%)					
	Tax Bill	Before Exemptions	Tax Bill	Before Exemptions	Tax Bill	Before Exemptions
	(1)	(2)	(3)	(4)	(5)	(6)
Black Mortgage Holder	14.8834***	12.2187***				
	(1.9459)	(2.0551)				
Black or Hispanic Mortgage Holder			11.3977***	8.0480***		
			(1.4335)	(1.5783)		
Other Nonwhite Mortgage Holder					3.2118***	2.0736***
					(0.2287)	(0.2737)
Jurisd-Year FE	Y	Y	Y	Y	Y	Y
Other Controls	N	N	N	N	N	N
No. Clusters	26371	26371	26371	26371	26371	26371
Observations	3,373,164	3,373,164	3,373,164	3,373,164	3,373,164	3,373,164
R^2	0.6803	0.6481	0.6802	0.6478	0.6802	0.6478
Note:						*p<0.1; **p<0.05; ***p<0.01

Note: This table repeats our baseline analysis in Table 1.1, but uses effective tax rate as the dependent variable instead of assessment ratio. Coefficients are percentages. For each racial and ethnic grouping, we present two sets of results. In odd columns, we show results using an effective rate computed using the observed tax bill and observed market value in the same year. Because the observed tax bill is potentially net of a wide range of exemptions, we also compute a before-exemption effective tax rate, by adding observed exemptions to the observed tax bill, and then dividing by market value. We trim any observation above a calculated effective tax rate of 25% both before and net of exemptions. We believe this to be a conservative choice as 25% is far higher than any property tax rate of which we are aware (the national median is approximately 1.4%), and is more likely than not to be a data error. All specifications use jurisdiction-year fixed effects to hold constant the level of intended taxation. Standard errors are clustered at the jurisdiction level.

Table 1.15: Racial Differential in Transacted Prices

	Proportional Realized Price Difference		
	(1)	(2)	(3)
Black Seller	0.022***		
	(0.002)		
Black or Hispanic Seller		0.033***	
		(0.002)	
Other Non-White Seller			0.005***
			(0.001)
Fixed Effects	Jurisd-B.G.-Yr	Jurisd-B.G.-Yr	Jurisd-B.G.-Yr
No. Clusters	18984	18984	18984
Observations	2,196,003	2,196,003	2,196,003
R^2	0.801	0.802	0.802
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table shows results from regressing the log difference of realized market price and predicted market price on a block-group-year fixed effect and categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect a racial differential in transaction prices net of predicted price. The predicted price is generated using zip-code level home price indexes. Standard errors are clustered at the jurisdiction level.

Table 1.16: Synthetic Assessments Using Zip Code HPIs

	log(Assessment) - log(Market)					
	Real Assessments			Synthetic Assessments		
	(1)	(2)	(3)	(4)	(5)	(6)
Black Mortgage Holder	0.144***			−0.041***		
	(0.015)			(0.003)		
Black or Hispanic Mortgage Holder		0.110***			−0.051***	
		(0.011)			(0.003)	
Other Nonwhite Mortgage Holder			0.031***			−0.007***
			(0.002)			(0.001)
Jurisd-Year FE	Y	Y	Y	Y	Y	Y
Other Controls	N	N	N	N	N	N
No. Clusters	18853	18853	18853	18853	18853	18853
Observations	2,135,943	2,135,943	2,135,943	2,135,943	2,135,943	2,135,943
R^2	0.910	0.910	0.910	0.712	0.713	0.713

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results from our proposed approach for correcting the assessment gap. Using the algorithm described in Section 1.6, we construct synthetic assessments using zip-code-level HPIs. We use Zillow's publicly available ZHVI series by zip-code. Our approach uses an initial transaction to pin down the base assessment value. At every subsequent transaction, we observe a realized assessment ratio along with our synthetically constructed assessment ratio. Columns 1–3 show that the overall assessment gap looks similar in the subset of homes to which can we apply this approach (smaller chiefly because the first transaction is not included in the analysis). Columns 4–6 show the assessment gap using our synthetic assessment ratios. All specifications include jurisdiction-year fixed effects. Standard errors are clustered at the jurisdiction level.

Chapter 2

Are Unfunded Public Pension Liabilities Capitalized in Local Real Estate Markets?

2.1 Introduction

State and local governments make large, inter-generational financial commitments to public employees in the form of defined benefit pension plans. Public pension funds, which administer these benefits, manage a pool of assets which are dedicated to covering future liabilities. When the current value of invested assets is equal to the discounted present value of earned retirement benefits, a pension fund is "fully funded". It has been widely documented that many public pension funds in the U.S. are significantly underfunded (Novy-Marx and Rauh, 2009, 2011, Rauh, 2016). Retirement benefits are contractually specified, and generally receive strong legal protections. As a result, unfunded public pension liabilities will require the government sponsor of the fund to make direct contributions to the pension system. Because the duration of the liabilities is measured in decades, governments facing an excessive shortfall typically plan to close the gap over a period of many years, rather than with a single lump sum payment – but nonetheless, governments sponsoring a fund with assets that cannot plausibly cover future liabilities from investment income must allocate revenue to their pension systems *at some point*. Any payment to a pension system entails a trade-off for the associated government: that payment can come from existing revenue streams or new ones. In the former case, the government's current operating budget will be reduced. Because state and local governments face balanced budget requirements, this implies a reduction in public goods or services. In the latter case, a new source of revenue means additional taxation, user fees, or public borrowing. Therefore unfunded public pension liabilities represent a future need to tax or to curtail services.

This paper uses local housing markets as a laboratory to test whether local residents seem to internalize the future need to tax – or to reduce services – represented by unfunded public pension liabilities. Specifically, I test whether a plausibly exogenous shock to unfunded public pension liabilities has predictive power for house prices. Two factors motivate the link between pension debts and housing markets. First, a long literature in urban economics posits a spatial equilibrium across regions, where regional amenities (dis-amenities) are balanced against higher (lower) costs in such a way that everyone receives the same utility (Tiebout, 1956, Rosen, 1974, Roback, 1982, Gyourko and Tracy, 1991). A shock to pension assets today implies some non-negative probability of reduced services or higher taxes in the future. Second, the central financial pillar for cities, counties, and other local governments is the property tax. When local governments need to raise revenue, this is one of the largest levers available to pull. Therefore, in a spatial equilibrium with fully rational homebuyers, an increase in future dis-amenities (either higher taxes or lower services) would translate to a lower willingness to bid for homes today.

My main empirical exercise tests whether changes in unfunded pension liabilities affect subsequent house price growth. I estimate a cross-sectional forecast regression of home price growth between 2010 and 2016 on changes in unfunded pension liabilities during the Great Recession (2008-2009). My measure of changes to unfunded liabilities comes from a novel panel dataset of financial flows for large pension funds which I obtain from hand-collected public accounting documents. Unfunded liabilities are aggregated to the county-level. I use two different measures of changes in home prices. First, I use an aggregated home price index for each county. Second, I use microdata of property-level transactions and focus on homes that sell in both 2010 and 2016.

I choose the Great Recession setting because pension funds experienced large losses in their investment portfolio during these years, and as a result I can observe large shifts in unfunded pension liabilities. Investment returns during this period are at least quasi-exogenous with respect to regional home price movements, as asset returns during 2008-2009 were primarily driven by macroeconomic shocks. However, there is still an endogeneity concern. Regions forecasting low economic growth might plausibly take on additional investment risk in an attempt to raise asset returns and thereby avoid having to send operating dollars to the retirement system in future years. In this case, asset losses during the Great Recession would be highest (and increases in unfunded liabilities largest) for those regions already predicting low future economic growth. In other words, changes in unfunded liabilities may well be endogenous with respect to persistent factors that affect regional home price growth.

To avoid introducing bias from this endogeneity, I construct an instrument for unfunded pension liabilities. I first obtain detail on fund asset class allocation. Then, using an asset pricing model, I compute the unexpected realized return to each asset class during 2008-2009. This unexpected return is the forecast error between the realized asset class return and the expected return based on asset class factor loadings in the Fama French 3-factor model (Fama and French, 1993). I argue that from the standpoint of regional economic conditions, unexpected return to broad asset classes should be considered as good as randomly assigned, making this a valid instrument.

My overall results suggest that unfunded pension liabilities do have a large effect on home price growth. This relationship, however, is little in evidence using county aggregates. With a county home price index as the dependent variable, I am unable to reject the null of no relationship in either the OLS or IV specifications. The strongest evidence, in fact, suggests a positive correlation between changes in unfunded liabilities and home prices – however, I show that this relationship becomes negative (though statistically insignificant) once controlling for public expenditures. While this is an endogenous control, the fact that signs flip upon its inclusion is noteworthy. This is consistent with the idea that financially constrained governments need to split a limited amount of public revenues between expenditures on goods or services (which residents value positively) and contributions to retirement systems; and that one potential cause of large pension debts is the decision to spend relatively more on services.

Using property-level data, I find much stronger evidence of a relationship between pension obligations and home prices. The OLS specifications show a strong negative correlation between changes in unfunded liabilities and subsequent home price growth. In response to a one standard deviation shift in per-capita unfunded liabilities, I estimate that a median-valued home would see a reduction in appreciation of $18,000 over the next 6 years. The IV specification suggests a similar relationship, however these findings are statistically insignificant. Property-level data also allows me to estimate effects by home value. I split my sample into high-value (above median) and low-value homes within each county. With this partitioning, I find an even stronger effect of unfunded liabilities on higher priced homes. These results are large and statistically significant in both the OLS and IV specifications. For an average high valued home, worth $424,000 in 2016, a one standard deviation shift in per-capita pension liabilities in 2008/2009 implies $75,000 less price appreciation over the next six years.

This research adds to the literature exploring the extent and effects of public pension under funding. Novy-Marx and Rauh (2009, 2011) and Brown and Wilcox (2009) document that unfunded public pension liabilities are large and under-reported relative to the (minimal) economic risk of future commitments. Rauh (2016) documents that annual financial flows to retirement systems are inadequate to close, or even hold static, funding gaps for the vast majority of states and large city governments. Myers (2019) models the relationship between risky pension debts, public services, and municipal insolvency. Using data from California, he estimates that stock market shocks to pension assets significantly increase the likelihood that municipal governments declare a fiscal emergency.[1] Boyer (2020) shows that higher unfunded liabilities within state pension plans subsequently increase state borrowing costs. The contribution of this paper is to show real economic effects of public pension underfunding. I show that unfunded pension liabilities directly affect the path of home prices. Homes are the most valuable asset for most households,[2] and therefore the pass-through from pension debts to real estate prices means that the financial structure of retirement systems is not simply an accounting nuance related to off-balance sheet debt of the municipal corporation, but rather a meaningful choice than can measurably impact household financial well-being.

The paper proceeds as follows. Section 2.2 presents background details on public pensions plans in the U.S. and describes the empirical strategy. Section 2.3 details the

[1]In California, such a declaration is often a necessary step in establishing eligibility to raise local tax rates under state legislation intended to limit such increases.

[2]2016 Survey of Income and Program Participation, 2014 Panel, Wave 4

data used. Section 2.4 presents the results. Section 2.5 concludes.

2.2 Setting and Empirical Strategy

2.2.1 Unfunded Public Pension Liabilities Represent A Cost to Local Taxpayers

Employer sponsored pension plans broadly fall into one of two categories: defined benefit (DB) or defined contribution (DC). In the former, payments made during retirement are set in advance according to a formula which generally depends heavily on years of service and salary level while working. In DC plans, retired employees are entitled to the individual contributions which have been made in their name plus investment proceeds. The distinction is where the risk lies: in DB plans, payments are fixed and therefore risk lies with the plan sponsor. For DC plans, the level of payments depends on investment returns achieved by each individual retiree account, and therefore the risk lies with the employee.

In the mid-20th century, the majority of workers in both the public and private sector were covered by defined benefit pension plans (Munnell et al., 2007). The last several decades have seen a large shift towards DC plans in the private sector, however DB plans have remained prevalent in the public sector. In 2019, 86% of state and local government employees had access to a defined benefit pension plan.[3] Every state in the US provides DB pension benefits to retirees, along with many large counties, cities and independent local governments (such as school districts).[4] In addition, larger local governments commonly sponsor their own DB pension plans. In 2017, the U.S. Census Annual Survey of Public Pensions reflects 433 county-level plans, 4,320 plans provided by cities and townships, 439 plans provided by autonomous quasi-governmental entities (predominantly municipal utility districts, transportation authorities, or fire districts) and 20 plans sponsored by independent school districts. Smaller counties and towns also provide DB pension benefits, but often delegate management to a state-level "multi-employer" plan rather than establishing a distinct pension fund. In 2017, there were 291 state-level DB pension plans across the U.S.[5]

A public employee earns DB retirement benefits during her years of service. The exact methodology for determining benefit levels is typically moderately complex and is often an outcome of collective bargaining between public sector unions and the employing government. An illustrative example would specify that any employee serving a minimum of 20 years would be entitled to receive annual benefits upon retirement equal to 80% of the average salary of her top three earning years. In addition, these benefits might be indexed to inflation in some way, and could include some provision for proportional reduction if the total years of service failed to exceed 20. While there are several differing accounting

[3] Bureau of Labor Statistics, National Compensation Survey- Benefits, data series NBU31900000000000028290.

[4] Two states, Alaska and Michigan, no longer provide any DB option for new employees (Munnell et al., 2007), however both states manage legacy DB systems with ongoing commitments to current retirees.

[5] Author's calculations using Annual Survey of Public Pensions 2017 data files.

methodologies for exact measurement of future payments (Novy-Marx and Rauh, 2011), the broader principle is simple: employees earn today benefits which will be paid in the future.

Pension fund liabilities are, therefore, financial commitments made by state and local governments. To cover these liabilities, pension funds manage a pool of invested assets. The source of these assets is contributions from employees (if these are required under the employment contract) and contributions from the plan sponsor. Public accounting standards require pension funds to make annual disclosures concerning the fiscal stability of the pension system.[6] As discussed below, there are accounting nuances in the measurement of both sides of a pension fund's balance sheet. This paper relies on the actuarial value of assets (AVA), along with the standard measure of liabilities reported in pension fund accounting documents. This measure, "Actuarial Accrued Liability" (AAL), is intended to capture the present discounted value of future benefit payments that have already been earned by employees.[7] The top line measure of DB pension fund fiscal stability is a funded ratio: if the fund's AVA is equal to it's AAL, a plan is 100% funded.

It is a stylized fact that most public pension funds in the U.S. are less than 100% funded. Of the 193 funds in my final sample, 187 had a funded ratio less than 1 as of 2010. The median funded ratio is 69%. Rauh (2016) detail similar under-funding in 2017 for 649 of the largest state, county, and city pension plans in the U.S. A gap between the current value of a fund's assets and its liabilities is referred to as unfunded liability - or, more precisely, "Unfunded Actuarial Accrued Liability" (UAAL).[8] Future cash flows to employees generally receive strong legal protections: in most U.S. states this takes the form either of explicit constitutional protections for pension benefits, or judicial recognition of retirement benefits as contractually protected obligations of the state or local government (Brown and Wilcox, 2009).

UAALs, therefore, are a form of public debt. From the standpoint of local residents, an increase of $1 UAAL represents an additional financial claim on the local government, for which tax payers are ultimately responsible. Although pension obligations are paid out over decades, meaning that necessary increases in taxation could be very far into the future, the principle of Ricardian equivalence holds that taxpayers will internalize today a reduction in wealth implied by any need to impose taxes in the future (Ricardo, 1824, Barro, 1974). The fact that pension liabilities are measured in present value makes the valuation of these future debts particularly simple. If all taxpayers have the same individual discount rate as that chosen by the local pension system, then local residents would value an additional dollar of UAAL as a (collective) cost of $1. Even without the restrictive assumption of homogeneous individual discount rates, the broader point holds: a positive shock to UAALs represents a strictly negative wealth shock to local residents in aggregate.

One technical note is important. There are large, meaningful nuances in the measurement of pension liabilities – most saliently the rate at which future benefit payments are

[6]This paper relies on disclosures mandated by Governmental Accounting Standards Board (GASB) Statements 67 and 68.

[7]In keeping with GASB Statement 67, the term "Actuarial Accrued Liability" denotes the same concept as "Total Pension Liability" - another commonly used term.

[8]A negative UAAL value would represent an excess of assets relative to liabilities.

discounted to the present. Funds are currently constrained to use their long-term expected rate of return on investments, or a high-quality municipal bond rate for highly underfunded plans.[9]. Many economists have argued that a lower discount rate is appropriate in order to match the essentially risk-free nature of the benefit payments (Lucas and Zeldes, 2006, Brown and Wilcox, 2009, Novy-Marx and Rauh, 2009, 2011). For purposes of this paper, I measure fund liabilities as stated in the fund's public accounting documents, without making any adjustment to harmonize discount rates. The empirical analysis rests on changes in unfunded liabilities driven by changes in asset values. In this first-difference specification, any measurement error on the liability side will be differenced out as long as that error is constant over a period of 1-2 years. As the level pension fund liabilities is an extremely slow-moving variable, especially relative to asset values during times of market volatility, I argue this simplification will not alter my results in any meaningful way.

2.2.2 Is Pension Debt Reflected In House Prices?

I use local housing markets as a laboratory for testing whether local residents are, in fact, sensitive to the costs embodied by pension UAALs. As detailed in Chapter 1 of this work, property taxes are the largest source of revenues for local governments. For the average local government, property taxes comprise 56% of the total revenue streams which local authorities can directly affect.[10] This makes it quite plausible that future revenue shortfalls will be closed by increasing property taxes. However, local governments certainly have a range of levers to pull in order to close fiscal shortfalls. This might include increasing user fees for any range of services provided by local governments (such as permits, inspections or licensing) or simply reducing spending in other categories. Of course all of these actions will also most directly affect local residents.

A central tenet of urban economics is the notion of spatial equilibrium, which holds that individuals are mobile and prices adjust until a constant level of utility is achieved across regions (Tiebout, 1956, Rosen, 1974, Roback, 1982, Gyourko and Tracy, 1991). In this framework, an exogenous increase in some future need to tax faced by one region would imply either a reduction of some cost or an increase of some non-pecuniary amenity. Amenities (typically viewed as things like weather or natural resources) are presumably fixed in short run. Given the central role of the property tax in local finances, this paper tests for a reduction in local housing market costs.

My central empirical exercise is a cross-sectional forecast regression of change in house prices on a shock to unfunded pension liabilities:

$$\Delta HP_{i,2016-2010} = \alpha + \beta \Delta UAAL_{i,2010-2007} + \epsilon_i \tag{2.1}$$

In this regression i denotes a county. The right-hand side variable denotes changes in county-level per-capita UAALs during the Great Recession. Changes in UAAL between 2007 and 2010 are aggregated across funds within a given county and then normalized by county-population. This transformation seeks to capture the portion of the region-wide

[9]GASB Statement 68.

[10]Author's calculation using Census of Governments data.

change in UAALs which would devolve upon one individual homeowner.[11] The dependent variable is some measure of changes in home prices between 2010 and 2016. My first set of results uses county-level home price indexes. I use indexes from both Zillow and the Federal Finance Housing Authority (FHFA). Results are not sensitive to choice of index. My second set of results uses observed property-level transaction prices directly, considering homes which sell in both 2010 and 2016.

The null hypothesis of $\beta = 0$ is that changes in unfunded pension liabilities have no ability to forecast changes in house prices. Any null result in equation 2.1 would not rule out that unfunded liabilities are very quickly capitalized into house prices. I show that evidence from a contemporaneous cross-sectional regression does not support this notion. One limitation of any null finding in this analysis is that I cannot rule out that pension debts do forecast house prices at a remove of more than 6 years.

Large investment shocks to funds assets during this period represent plausibly exogenous shifts in UAALs. The financial crisis in 2008-2009 resulted in large losses across many asset classes. Pension funds, which often seek passive exposure to a broad market portfolio, saw large losses during this period. Across the 105 counties analyzed in this paper, 102 counties realized net losses during 2008 and 2009. In 2008, 96% of individual funds lost money.[12]

Although negative returns on invested assets during 2008 and 2009 seem most likely related to the macroeconomic shock of the Great Recession rather than local economic conditions (which one would expect to correlate with local home prices), fund returns may still be endogenous with respect to house prices. Regions facing economic challenges may have underfunded pension plans as a result of financial constraints. If these regions seek to make up the shortfall by taking additional risk exposure with invested pension assets, then negative returns during a broad market downtown would positively covary with local economic stress. The latter would presumably be reflected in house prices through a range of channels.

To address this endogeneity, I construct an instrument for changes in UAALs. I obtain data on funds' asset allocation from the US Census of Governments. I then use an asset pricing model to predict expected returns for each asset class. For each asset class, I take the unexpected return to each asset class – the forecast error – and aggregate to the fund level using the reported asset class weights. I use this total unexpected return as an instrumental variable in a two-stage least squares regression:

$$\Delta HP_{i,2016-2010} = \alpha + \beta \Delta \widehat{UAAL}_{i,2010-2007} + \epsilon_i \tag{2.2}$$

$$\Delta \widehat{UAAL}_{i,2010-2007} = \psi + \Theta unexpected\_return_{i,2009} + \delta unexpected\_return_{i,2008} + \nu_i. \tag{2.3}$$

Unexpected return is computed for each fund and then aggregated to the county level as:

[11] I do not adjust this per-capita measure to reflect the fact that average household size is greater than one individual.

[12] Author's calculation using data assembled from fund financial disclosures.

$$unexpected\_return_{it} = MVA_{i,t-1} \sum_j w_{ijt}(r_{jt} - E[r_{jt}]) \tag{2.4}$$

where j index asset classes. Note that the realized returns are not indexed by either fund or county; these are realized returns for the asset class benchmark. Expected return for each asset class is calculated using the Fama-French 3-factor asset pricing model, augmented with LIBOR (Fama and French, 1993):

$$E[r_{jt}] = \hat{\beta}^j_L LIBOR_t + \hat{\beta}^j_M (r_m - r_f) + \hat{\beta}^j_{SMB} SMB_t + \hat{\beta}^j_{HML} HML_t + e_{jt}. \tag{2.5}$$

Asset class betas are estimated from historical returns of asset class benchmarks.

The logic of the instrument is this: in any given year, assets classes as a whole may under- or over-perform relative to the asset-pricing model forecast. As long as these unexpected returns are orthogonal to regional drivers of home prices, the exclusion restriction will be satisfied. Given this, there are two ways of motivating instrument validity. The first is the standard asset pricing result that forecast errors in an asset pricing model are unpredictable (Froot and Frankel, 1989). If any regional factor could predict the forecast error in asset class performance, then in an efficient market, this would be incorporated into asset prices ex-ante. Another motivation recognizes this instrument is effectively constructed as a shift-share instrument: the local vector of asset class weights is aggregated using national shocks. These weights are analogous to base-shares in a classic Bartik instrument, and the unexpected asset class return would be correspond to industry sector shocks (Bartik, 1991). Because the weights do not derive from a pre-period, they cannot be considered exogenous. However as Borusyak et al. (2018) show, base shares do not need to be exogenous as long as the shocks are numerous and sufficiently orthogonal to factors affecting the dependent variable (the condition is slightly weaker than full exogeneity). Viewed through this lens, the identification assumption in this setting becomes that asset class performance relative to the factor model prediction is exogenous with respect to regional economic factors. Given that these broad asset classes represent national or international investments, this seems a relatively mild assumption.

2.3 Data

2.3.1 Novel Pension Dataset

The core data of this paper is a novel panel dataset of unfunded liabilities and investment returns for just over 200 of the largest local DB pension plans in the county during 2006-2016. I extract this information by hand-collecting public accounting documents for each fund. There is a relative lack of any compiled dataset on local pension funds. The standard sources of data on public pensions are the Census Bureau's Annual Survey of Public Pensions and Government Finance Statistics. Neither of these surveys, however, asked funds to disclose liabilities until the 2017 vintage. As a result, it is not possible to ascertain a measure of unfunded liabilities from this data. To the best of my knowledge, the data used in this paper represents the broadest snapshot of local pension fund finances that has been compiled. There are other data which have partial overlap with my dataset. The Center for Retirement Research (CRR) at Boston College maintains a database spanning 190 large pension plans and extending back as far as 2001. At the time I collected data for this paper, the CRR resource focused primarily on states. In recent years, that database has expanded to include some large counties and cities. Joshua Rauh has also compiled more recent data beginning in 2016 on pension liabilities, which is made available through a Hoover Center publication entitled *Hidden Debts, Hidden Deficits* (Rauh, 2016).

Local funds are those sponsored by sub-state governments: counties, cities, towns, and other regional public sponsors like independent school districts. I focus on local funds because it is these governments that derive a significant portion of their revenue through property taxes. Property taxes are not typically important for states, comprising less than 2% of total revenue in recent years.[13] By restricting attention to local funds, I am focusing on pension debt that is most likely to impact home prices.

The data itself comes from disclosures mandated by the Government Accounting Standard Board Statement 67 and 68. These statements, issued in 2014, altered existing guidelines to expand the reporting detail required, most saliently around fund discount rates and the components of changes in Net Pension Liability (equivalent to UAAL). Pension funds disclose this information annually in a Comprehensive Annual Financial Report (CAFR).[14,15] GASB 67 mandated disclosure of ten years of historical information, which greatly enabled the compilation of the panel.

I first use the Census Bureau's Annual Survey of Public Pensions to identify the largest pension funds in the U.S. by assets.[16] Then, I hand-collect CAFRs for each fund and use these to build a panel of the following variables over the 2006-2016 period: market value of assets, actuarial value of assets, actuarial accrued liabilities, investment return, net investment income, required employer contribution, actual employer contribution, member contribution, and fund discount rate. My measure of UAAL is actuarial accrued liabilities (AAL) minus actuarial value of assets (AVA).[17]

It is relatively common for a city or county to sponsor several pensions funds (for instance, separate funds for teachers, police and fire, and general public administrators). Since my baseline analysis is conducted at the county level, whenever I come across multiple funds sponsored by the same government, I collect information for those fund as well, even if they do not appear on the list of largest funds by assets.

I aggregate UAALs across all funds within a county for each year, and then divide

---

[13]Author's calculation using Census of Governments Government Finance Statistics, 2013-2016

[14]A CAFR is the public sector equivalent of a 10-K filing, and GASB is the public sector counterpart to the Financial Accounting Standards Board. GASB is an independent non-profit organization and does not have regulatory enforcement authority. Nonetheless, public entities follow these standards very closely.

[15]While most pension funds produce a separate CAFR, sometimes the information is disclosed in the CAFR of the sponsor government instead.

[16]The optimal list would include the largest funds by unfunded liabilities, however it is exactly this lack of data that the overall exercise seeks to address.

[17]Relative to market value of assets (MVA), the major feature of AVA is smoothing of investment returns. Broadly, results are similar defining UAAL = AAL - MVA. The first stage relationship in the IV regressions is stronger, which is a mechanical result of removing the measurement error introduced by asset smoothing.

by county population to get my baseline measure of per-capita unfunded liabilities. For my baseline specification, I use the changes in per-capita UAALs between 2010 and 2007. Table 2.1 shows summary statistics for county-aggregate public pension financials as of 2010.

2.3.2 Other Data Used

In order to construct the instrument, I obtain information on fund asset allocation from the US Census of Governments, along with detail on total expenditures and revenues for sponsor governments. Fund holdings are classified into: cash, savings, US Treasuries, agency debt, state and local debt, corporate stocks, corporate bonds, mortgages, foreign stocks, real property, savings, and miscellaneous. For each asset class, I select a passive benchmark for which I can observe monthly returns. From The Center for Research in Security Prices (CRSP), I obtain the historical time-series for each benchmark extending back to 2005, and use these to compute monthly betas for each asset class between 2007 and 2009. (I compute rolling betas updated monthly using the return series from 2005 to the current month). Table 2.2 shows the list of asset classes along with the passive vehicle selected to represent that asset class.

For a measure of home prices, I use county-level Home Price Indexes (HPIs) from both Zillow and the Federal Housing Finance Authority. I obtain county-to-county population inflows and outflows between 2010 and 2014 from the American Community Survey. From the Census of Governments Government Finance Statistics, I obtain several fiscal aggregates for local governments: total revenues, general revenues (which excludes intergovernmental-transfers), property tax receipts, total expenditures, and direct expenditures (this excludes capital spending).

I obtain real estate transactions records from ATTOM. This is a comprehensive dataset of 53 million transactions between 1999 and 2016. Transaction data is sourced from county recorder offices. Each property is characterized by a unique identifying ID, which allows me to focus on repeated transactions. The recorder portion of the ATTOM dataset has several indicator flags for arm's-length transactions and partial interest sales, which collectively can be used to exclude transactions that may not provide an accurate signal of market value. I include only homes which sell in an arm's-length, full consideration transaction. Each property-year observation also includes a large vector of property characteristics, including size, number of rooms, age, and several binary indicators for various property features.

2.4 Results

I present results in two sections: 1) results using county-level aggregates, and 2) evidence using individual level home prices.

In the first section, I use a county-level home price index as the dependent variable, and estimate forecast regressions of index changes between 2010 and 2016 on changes in unfunded liabilities between 2007 and 2010. At this level of aggregation, I find very little evidence that unfunded pensions liabilities affect home prices. Without a control for the 2010 level of public expenditures, the data suggests a positive relationship between unfunded liabilities and home price growth. The inclusion of this control for public spending causes the point estimates on unfunded liabilities to switch sign and become negative. Estimates lack precision, however, and I cannot reject the null of no impact.

In the 2nd section, I use property-level data from real estate transactions, and find much stronger evidence of a relationship between home prices and pension liabilities. I find a large, and highly significant, negative relationship between shocks to pension liabilities and future home price growth. In addition, having multiple data points per county allows me to test for differing effects by property value. I find strong evidence that pension liabilities have a larger impact on price growth for more valuable properties. This is consistent with either a belief that taxation (or service curtailment) will ultimately fall more heavily on higher-priced homes, or with a greater internalization of pension debts by wealthier homeowners, who may be more financially sophisticated.

Unless otherwise noted, all estimates presented in Tables 3–12 are multiplied by 100.

2.4.1 County Level Aggregate Results

Table 2.3 shows the results from an OLS estimation of equation 2.1. This is a forecast regression that tests whether changes in unfunded public pension liabilities can predict changes in home prices. I regress the six-year log difference in a county-level home price index (HPI) on changes in UAALs that occur during the Great Recession. I present results considering changes in UAALs from the end of 2007 to the end of 2010, and also results using changes in UAALs over a shorter, two-year period between 2007 and 2009. Results are not sensitive to the period over which UAALs are measured. My preferred specification uses the change between 2007 and 2010. Public entities report data with a variable lag and at different times during a year (driven, often, by different choices of ending month for the fiscal year), and so the longer window of time seeks to capture the total change in unfunded liabilities arising from the Great Recession. I also present results using two different measures of home prices: the county-level HPI produced by the Federal Housing Finance Authority, and one produced by Zillow. The two are, unsurprisingly, very highly correlated: the correlation is 95.4% over the 2010–2016 period.

For ease of consumption, all figures in Table 2.3 are multiplied by 100. These results suggest that an increase in unfunded liabilities correlates to *higher* future home price growth. The estimated coefficient in column (1) suggests that a county-wide increase of \$1 in UAAL would correspond to in increase in home price growth of .00003 percentage points. The sample standard deviation of $\Delta(UAAL)$ between 2007 and 2010 is just over \$1,000. The average county in the sample has a population of 1.02 million, and so a one-standard deviation shock in per-capita UAALs would imply a shift in unfunded liabilities of \$1.02 billion. The average of county-level pension assets in this sample is \$3.69B. If this hypothetical shift in UAAL were driven entirely by investment losses, this would be an investment return of -27.8%. This is certainly a large shock, but not unrealistic given asset price movements during the Great Recession. The peak to trough decline in the S&P500

during this period was -53%.[18] Column (1) implies that a large, but plausible, one-standard deviation increase of $1,000 per-capita would result in an additional 2.77 percentage points of growth in home prices between 2010 and 2016. For a $200,000 home, this would be an additional $5,500.

This finding is counterintuitive. As noted, there are reasons to be concerned about omitted variable bias in the estimates of Table 2.3. I turn to this issue next. Before proceeding, however, I highlight another interpretation that would explain this surprising finding. As described in Section 2.2, unfunded public pension liabilities are off-balance sheet debt for the sponsor government. Local governments are subject to balanced budget requirements. If governments effectively borrow from their pension systems in order to spend more on current services, and if these services are more salient to local residents than pension debt, then higher UAALs would result in higher home prices. As described, that reaction from local residents implies a failure of Ricardian Equivalence: residents would place a positive value on the good and services provided with borrowed money, but would fail to internalize the cost of repaying that debt in the future. The remaining discussion in this section attempts to distinguish between these two conflicting interpretations.

First, I estimate the IV version of the baseline analysis using equations 2.2 and 2.3. This will remove any omitted variable bias in the previous estimates. Table 2.4 shows the first stage results from 2SLS estimation. I construct unexpected return for both 2008 and 2009, but find the strongest first stage relationship when I use unexpected return from 2008 alone. Following Stock and Yogo (2002), I report the first-stage partial F-statistic. The first stage is strong when unexpected return instruments the change in UAALs between 2007 and 2010. The F-statistic is just under the rule-of-thumb benchmark of 10, even using heteroskedastic-robust standard errors to compute it. With the assumption of homoskedasticity, the first stage F-statistic is 88.[19] The first stage is weaker when unexpected returns are used as an instrument for changes in UAALs between 2007 and 2009. I present results using both measures of changes in UAALs. The typical concern about weak instruments in a just-identified 2SLS setting is not about bias, but rather that an estimated coefficient will be too large, yielding a spurious rejection of the null (Angrist and Pischke, 2008). My IV estimates will not, in fact, end up rejecting the null and so I am less concerned here about a weak first stage than I would be otherwise.

Table 2.5 shows the results from the second stage estimation. Most estimates are slightly smaller in magnitude and all are statistically insignificant. I cannot reject the null that changes in unfunded pension liabilities have no ability to forecast home price growth. The point estimates are still positive, and collectively imply that an increase in per-capita unfunded liabilities of $1,000 would result in additional $2–7000 in home value between 2010 and 2016.

The next set of results extends my baseline analysis by adding one control: the per-

[18] Author's calculation using data from CRSP.

[19] Staiger and Stock (1994) is typically viewed as the initial publication proposing that first-stage F-statistics should exceed 10 for an IV regression, however it is clear from the paper that this was already common practice at the time. The analysis in that paper is predicated on an assumption of homoskedastic standard errors.

capita total public expenditures in 2010.[20] Two main factors drive short-term changes in unfunded pension liabilities: shocks to the stock of invested assets, and net flows into the fund.[21] I choose the setting of the Great Recession precisely because of large quasi-exogenous shocks to asset prices occurring during this period. However, governments, of course, adjust their budgets in response to crisis. If public entities face any financing constraints, then a choice to send \$1 to retirement systems would represent an implicit choice not to spend \$1 on current operating expenditures.[22]

Formally, the level of expenditures in 2010 represents an endogenous control with respect to changes in unfunded liabilities during 2008-2009. However, the inclusion of this control allows me to distinguish between areas where unfunded liabilities have (potentially) risen because flows to retirement systems were curtailed to keep operating expenditures high, and areas where unfunded liabilities have (potentially) risen for reasons unrelated to changes in public goods provision. As Novy-Marx and Rauh (2009) point out, public pension liabilities are a particularly opaque form of off-balance sheet debt. If residents are not sensitive to pension debt, but do positively value services which are provided with dollars that would otherwise flow to the pension system, then an increase in pension liabilities would, in fact, correspond with an increase in home prices.

Tables 2.6 and 2.7 show the results from controlling for expenditures. In Table 2.6, the point estimates from the OLS regression are all negative. These estimates are not precise; none are statistically distinguishable from zero. It is suggestive, however, that the inclusion of the endogenous control for expenditures reverses the sign of the counterintuitive results from Table 2.3. And, unsurprisingly, the coefficient on per-capita expenditures is large and significant. The IV regression yields similar results. The signs again are all negative. The 1st stage F-statistic is lower when expenditures are included as a control. All estimates are statistically indistinguishable from zero. Although the results from Tables 2.6 and 2.7 lack statistical precision, the overall pattern is not only suggestive, but also matches the evidence of the next section when property-level prices are used: conditional on realized levels of public spending, unfunded pension liabilities appear to correlate negatively with home price growth, but unconditionally the sign of the relationship is less certain.

In my last set of results using county-level data, I test for a contemporaneous relationship between unfunded liabilities and home prices. The implicit assumption behind a forecast specification is that changes in unfunded pension debt are slow to enter prices. My analysis tests whether changes in net liabilities appear in house prices within a period of 8 years. However it is also possible that changes in pension debts are instantly reflected in housing markets. Were that the case, I would also find a null result in the forecast regressions. Table 2.8 shows the results of regressing changes in house prices during the

[20] I aggregate expenditures across all public entities within a given county using Census of Governments data.

[21] In the longer run, the level of pension benefits and numbers of public employees hired will have large impacts on fund liabilities. Liabilities evolve slowly, however, especially relative to shocks which hit the asset side of a fund's balance sheet.

[22] In separate work, I estimate public budget share elasticities with respect to a large fiscal shock during the Great Recession, and find that flows to retirement systems have a large negative elasticity (Howard and Morse, 2020).

peak of the Great Recession on changes in UAALs during that same period. Columns (1) and (2) show OLS results excluding and including a control for expenditures. Columns (3) and (4) show IV results for the same specification. I do not find a significant relationship in any column. The point estimates are also substantially smaller than the estimates in the forecast regression. Accordingly, I consider this to be a meaningful lack of evidence for any contemporaneous capitalization of pension liabilities into house prices.

2.4.2 Results using Property-Level Transaction Prices

One challenge of the preceding analysis is the small sample size. Because UAAL data is aggregated to the county level and a single HPI is used as the dependent variable, the regression includes only 105 data points. In this section, I repeat my analysis using property-level data on home prices. This offers several advantages. First, the number of observations expands dramatically. I use detail on real-estate transactions from the ATTOM database. This data is compiled from administrative records and represents near universal coverage of home sales. To parallel the analysis of the preceding section as closely as possible, I exact properties which sell in both 2010 and 2016, and use the log difference in prices as my independent variable. This yields 15,000 properties.

With multiple data points per county, I can add additional controls for regional economic factors which are either infeasible or impractical in the county-level regression. Specifically, I add state fixed effects and tract-level measures of economic vitality. The ATTOM data also includes property-level characteristics at the time of sale, and so I include a vector of first-differenced home attributes to ensure that changes in transaction price are not being driven by major renovations. In addition, the first-difference specification removes any (static) unobservable property-level features, along with any (static) regional features which might be a source of bias. The regressor of interest, $\Delta UAAL$, still does not vary within county. As this is likely to generate within-county correlation of residuals, I cluster standard errors at the county level in all specifications.

The restriction to homes which sell in both 2016 and 2010 is not without consequence: of the 105 counties for which I collect data on pension fund liabilities, I observe the requisite pair of transactions in only 48 counties. These 48 counties are slightly smaller the the excluded counties: the median county population in 2010 is 720,000 for included counties, and 801,000 for excluded counties. Homes within the covered 48 counties are also slightly more expensive: the average 2016 purchase price for homes entering the sample is \$305,000 in 2016 (\$240,000 for the median). The average 2016 purchase price for other homes which transact in both 2010 and 2016 but do not lie within one of these 48 counties is \$264,000 (\$200,000 for the median). Finally, pension liabilities are substantially lower within the in-sample counties: the median per-capita UAAL is \$373, compared to \$648. Within both included and excluded counties, the standard deviations of per-capita UAALs is similar at approximately \$1000. In this analysis I do not explore heterogeneous effects by total level of indebtedness. If that distinction is important, then the restriction to a set of relatively less-indebted counties is obviously important for the external validity of my findings.

Table 2.9 shows the results from estimating several versions of the following equation:

$$\Delta log(p_{i,c,2010-16}) = \alpha_{state} + \Delta UAAL_{c,2007-10} + E_{c,2010} + \Theta\Delta X_{i,2010-16} + \Phi\Delta W_{i,2010-16} + \epsilon_{i,c} \quad (2.6)$$

where i denotes a home, c a county, $\Delta log(p_{i,2010-16})$ is the log difference in transaction prices, $\Delta UAAL_{c,2007-10}$ is change in per-capita UAAL between 2007 and 2010, $E_{c,2010}$ is total public expenditures in 2010 in county c, $\Delta X_{i,2010-16}$ is a vector of 1st differenced geographic controls associated with a given property (varying at the census tract level), and $\Delta W_{i,2010-16}$ is a vector of 1st differenced property-attribute controls.

In all specifications, the coefficient of interest is negative. The estimate in column (2) which controls for expenditures but no other changes in property or geographic features suggests that a \$1000 per-capita increase in UAAL would result in 7.86 lower percentage points of home price growth. For a median-valued home of \$240,000, this suggests reduced appreciation of \$18,850 over 6 years. With the full set of controls, the estimate in column (4) suggests a reduction in value of \$18,000. All estimates are significant at the 10% level.

Table 2.10 shows the results from the IV specification of this regression. Here, all estimates are statistically insignificant. The same pattern observed in county aggregates is present here as well: point estimates are positive without controlling for expenditures and negative with the inclusion of the spending variable. The 1st stage F-statistics are again low enough that some concern about weak instrument bias is warranted – although, as in the previous section, all F-statistics are much higher than 10 under the assumption of homoskedasticty. The low number of clusters here poses a challenge for statistical power in both the 1st and 2nd stage.

The use of property-level data also permits me to test for different effects by property value. If homeowners believe that future revenues will be raised through an increased property tax levy, it is possible that owners of higher valued homes might react more strongly. Certainly this would be a reasonable expectation if the expected additional levy were substantially progressive. The use of "mansion taxes" to address pubic revenue shortfalls has been an ongoing part of the political discourse in recent years, and so this is not a farfetched possibility.[23]

Within each county, I classify homes as being "high value" or "low value" based on whether the 2016 transaction price is above or below the sample median. I then repeat the preceding regression with this partitioning for the regressor of interest. Table 2.11 shows results from the OLS regressions. The point estimates for high-value homes are all negative and strongly statistically significant. The point estimates for low-value homes are small, positive, and statistically indistinguishable from zero. The estimated magnitude of the effect on high valued homes is economically very large: the coefficient in column (2) suggests that a \$1,000 shift in UAAL would result in 20 log points lower growth. For the average "high valued" home, which transacted at \$424,000 in 2016, this would represent a reduced price of nearly \$93,000. With full controls, the estimate from column (4) suggests a response of \$89,500. While these estimates are quite large, it is important to remember that a shock of an additional \$1,000 in unfunded liabilities would represent a near-quadrupling

[23]See, for instance, "NYC Brokers Relieved as Mansion Tax Replaces a Pied-a-Terre Levy", Bloomberg News, April 1, 2019.

of off-balance sheet debt for the median county. In this context, a large economic reaction housing markets is not *a priori* unreasonable.

Again, I repeat the analysis instrumenting for changes in UAAL. Table 2.12 shows the results. Here, the IV estimates are very consistent with the OLS evidence. Point estimates are again negative for high-valued homes. Magnitudes are slightly smaller than in the OLS regressions. Again relative to an average high-valued property transacting for $424,000 in 2016, the estimates of columns (2) and (4) suggest a response of $75,000 and $72,700 respectively. The estimated coefficient on high-valued homes in columns (2) and (4) are significant only at the 10% level. The corresponding estimate is not significant without the expenditures control in columns (1) and (3). For most specifications, the estimated coefficient on low-valued homes is not significant. It is also positive across all columns. This is consistent with owners (and buyers) of lower-priced homes valuing whatever is provided with funds that otherwise would have gone to retirement systems more than the disamenity value of additional public debt. Given the lack of statistical significance, I do not argue too strongly for this interpretation, but consider it to be a fruitful subject for further research. In all cases, the estimated coefficient for high-valued homes is strongly statistically different from that for low-valued homes. This strongly supports the notion of heterogeneity in reaction to public debt by wealth levels. This, also, is a promising area for further analysis.

2.5 Conclusion

I test whether unfunded public pension liabilities are capitalized into house prices. Using data from public accounting documents, I assemble a novel panel data set of annual financial flows for more than 200 of the largest defined benefit public pension plans in the United States. I observe large quasi-exogenous shocks to fund investments during the Great Recession, which induce changes in per-capita unfunded liabilities. Given that individual taxpayers are the ultimate source of public revenues, and that local governments rely very heavily on property tax receipts, I argue that a rational homeowner would internalize increased expected future tax payments (or future curtailment of services) into home valuation. I show that changes in unfunded pension liabilities are not contemporaneously capitalized into home prices, and therefore focus the analysis on forecast regression of changes in home price between 2010 and 2016 regressed on changes in UAAL between 2007 and 2010.

I estimate several versions of this regression using both OLS and two-stage least squares. The IV specifications rest on an instrument for changes in unfunded liabilities between 2007–2010 which I construct from pension fund asset allocations and investment returns during 2008 and 2009. Any correlation between future home price movements and contemporaneous pension fund returns represents a source of endogeneity bias. In particular, if a region forecasts economic difficulty that will depress both home prices and public revenues, it is possible the pension fund would deliberately seek more risk in its investment allocations. During the Great Recession, then, large losses would positively covary with lower home price growth.

To address this, I instrument changes in UAAL between 2007–2010 with asset-class idiosyncratic return. I first obtain fund disclosures of asset class allocation. Then, for every asset class, I select a passive benchmark. I use an asset pricing model to construct factor loadings for each asset class, and then in conjunction with realized returns to these factors, I compute the expected return to each asset class, conditional on market-wide performance. I use as my instrument the weighted sum of residuals across asset classes. During 2008–2009, asset classes performed slightly better or worse than the standard asset pricing model would have predicted. As long as these broad asset-class residuals are uncorrelated with local conditions, this instrument satisfies the exclusion restriction.

Using county-level home-price indices as the dependent variable, I find little evidence that pension debts affect home prices in either the OLS or IV specifications. At this level of aggregation, most of the estimated effects are statistically insignificant. Although unable to reject the null of no effect, I find point estimates suggesting a counterintuitive positive correlation between changes in pension debt and future home prices. Because pension plans receive annual inflows from the sponsor government, a change in unfunded pension debt is linked to decisions about spending – especially in the presence of any financial constraints. I argue that the inclusion of an endogenous control for total public expenditures is important to understand the sign of the relationship between changes in pension debt and changes in house prices. I find that with the inclusion of this expenditure control, the estimated relationship between pension debts and home price growth is consistently negative.

I then use property-level data on transacted home prices to repeat the analysis. From a dataset covering the near-universe of home transactions, I isolate properties which transact in both 2010 and 2016. I then take the log difference of transaction prices as my independent variable. Having multiple observations within a given county allows me to add more precise controls and also increases statistical power. In OLS regressions I find a negative relationship between UAAL and home price growth which is significant at the 10% level. The IV version of this analysis is unable to reject the null of no relationship, however the point estimates again are positive without a control for expenditures, and negative with the inclusion thereof, paralleling the county-level results.

Using the property-level data, I also test for evidence of heterogenous effects by property value. I split the sample into "high" and "low" valued homes – properties above and below the within-county mean transaction price. Changes in UAALs appear to have a much stronger affect on price growth for high valued properties. With this sample split, I find a strongly statistically significant negative relationship between UAALs and home prices for high-valued properties. This holds also with the IV, although statistical significance is only at the 10% level. The magnitude of the response for high-valued properties is large: for an average home in this sample of \$424,000, my estimates predict that a one standard deviation change in UAALs will result in \$75,000 reduced appreciation.

The totality of the evidence suggests that off-balance sheet debt represented by unfunded public pension liabilities does affect local real estate markets. The effects appear to differ by home value. It seems also very plausible that there is regional heterogenity based on existing public financial conditions. The potential for different responses by either existing level of public debt or by other measures of fiscal capacity are anticipated future areas of research. When effects in real estate markets are evident, they also seem to be large relative to the magnitude of changes in debt. Given the extremely large total stock of unfunded pension liabilities, and the central role of housing wealth in household finance, understanding these relationships better will help inform difficult public finance trade-offs that face retirement systems during any period of fiscal stress.

Tables

Table 2.1: Summary Statistics, County Level Pension Finances

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
AVA	105	4,187	12,676	17	525	2,974	105,269
MVA	105	3,669	10,948	16	511	2,708	91,229
AAL	105	5,776	18,780	236	788	4,348	165,203
UAAL	105	1,589	6,452	−495	91	925	59,935
UAAL per-capita	105	910	1,197	−817	156	1,160	7,316
Net Inv. Income, 2008	105	−433	1,017	−7,185	−333	−55	91
Net Inv. Income, 2009	105	−474	2,429	−20,425	−221	78	1,109
Changes, per-capita UAAL 2007-2010	105	578	1,002	−536	126	717	8,556
Changes, per-capita UAAL, 2007-2009	105	439	737	−351	95	561	6,393

All figures, expect per-capita measures, are in millions of dollars

Note: This table shows summary statistics, as of 2010, for measures of defined benefit public pension plan assets, liabilities and investment returns, aggregated to the county level. AVA, MVA, AAL, UAAL, and UAAL per-capita are as of 2010. UAAL is defined as AAL - AVA. The underlying financial data is extracted from Comprehensive Annual Financial Reports (CAFRs) for each pension fund separately. In some cases, figures are extracted from the CAFR for the parent government of a given pension fund sponsor.

Table 2.2: Asset Classes and Benchmarks

Asset Class (from Fund Allocation)	Passive Benchmark	Benchmark Ticker or Series No.
US Treasury	Vanguard Intermediate Treasury	VFIUX
Savings	FRED 3-month or 90-day US CD Rates	IR3TCD01USM156N
Fed Agency Debt	Vanguard GNMA Fund Admiral	VFIJX
State-Local Debt	S&P National AMT-Free Municipal Bond Index	n/a
Corporate Bonds	iShares iBoxx Investment Grade Corporate Bond ETF	LQD
Corporate Stocks	Vanguard Total Stock Market Index Fund	VTSAX
Mortgages	Pimco Mortgage-Backed Securities Fund	PTRIX
Foreign Stocks	Vanguard Total International Stock Index Fund	VGTSX
Real Property	Vanguard REIT Index Admiral Fund	VGSLX
Miscellaneous	Cambridge Associates LLC US Private Equity Index	n/a

Note: This table lists the pension fund asset classes categories from Census of Governments data (left-most column), along with the passive benchmark selected to represent the asset class. When applicable, the ticker for the passive benchmark is listed as well. The S&P National AMT-Free Municipal Bond index is not a traded vehicle and does not have a ticker. A private equity index produced by Cambridge Associates is used for the miscellaneous category; this also does not have a ticker. The benchmark for savings is a data series of 3-month CD rates obtained from FRED at the St. Louis Federal Reserve Bank. All other benchmarks are traded and the data on returns is obtained from CRSP.

Table 2.3: Predictive Regression of Home Price Growth on UAALs

	OLS Regressions			
	FHFA	Zillow	FHFA	Zillow
	(1)	(2)	(3)	(4)
Delta(UAAL per capita), 2007-2010	0.0028***	0.0019*		
	(0.0009)	(0.0010)		
Delta(UAAL per capita), 2007-2009			0.0031*	0.0020
			(0.0017)	(0.0017)
Constant	16.2047***	24.9915***	16.4299***	25.2233***
	(1.5925)	(1.9579)	(1.5922)	(1.9606)
Observations	105	105	105	105
R^2	0.0327	0.0111	0.0226	0.0065
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results for the baseline specification of regressing changes in home prices between 2010 and 2016 on changes in unfunded public pension liabilities. All estimates are multiplied by 100. The unit of observation is a county. All models are estimated with OLS. Columns (1) and (2) use the changes in UAALs between 2007 and 2010 as the independent variable. Columns (3) and (4) use only the change in unfunded liabilities between 2007 and 2009. The dependent variable is the log growth of a county-level home price index (HPI). Columns (1) and (3) use the Federal Housing Finance Authority HPI for the dependent variable, and columns (2) and (4) use Zillow's ZHVI HPI. All specifications report heteroskedasticity-robust standard errors.

Table 2.4: First Stage Results for Investment IV

	First Stage: Dep. Var is Delta(UAAL)	
	2007-2010	2007-2009
	(1)	(2)
Unexpected Return, 2008	4,199.00***	2,876.68**
	(1,388.40)	(1,192.99)
Constant	597.96***	452.90***
	(73.65)	(57.24)
Partial F-Stat	9.15	5.81
Observations	105	105
R^2	0.46	0.40
Note:		*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the first stage results from the IV regression of changes in home prices on changes in unfunded pension liabilities. The instrument is constructed from unexpected returns to asset classes during 2008, as described in Section 2.2. The left-hand side variable is changes in unfunded liabilities: between 2007 and 2010 in column (1) and between 2007 and 2009 in column (2). Both specifications report heteroskedasticity-robust standard errors. The reported F-statistic is for the instrument only and is computed using heteroskedasticity-robust standard errors.

Table 2.5: IV Regression of Home Price Growth on UAALs

	2SLS Regressions			
	FHFA	Zillow	FHFA	Zillow
	(1)	(2)	(3)	(4)
Delta(UAAL per capita), 2007-2010	0.0025	0.0010		
	(0.0018)	(0.0021)		
Delta(UAAL per capita), 2007-2009			0.0036	0.0014
			(0.0025)	(0.0031)
Constant	16.3635***	25.5522***	16.2060***	25.4913***
	(1.8317)	(2.2377)	(1.9071)	(2.3159)
1st Stage Partial F-Stat	9.15	9.15	5.81	5.81
Observations	105	105	105	105
R^2	0.0324	0.0083	0.0220	0.0059

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results for the baseline specification of regressing changes in home prices between 2010 and 2016 on changes in unfunded public pension liabilities. All estimates are multiplied by 100. The unit of observation is a county. All models are estimated using 2SLS. Each column in Table 2.5 is the IV counterpart of the OLS estimates presented in Table 2.3. The dependent variable is the log growth of a county-level home price index (HPI). Columns (1) and (3) use the Federal Housing Finance Authority HPI for the dependent variable, and columns (2) and (4) use Zillow's ZHVI HPI. All specifications report heteroskedasticity-robust standard errors. The partial F-statistic from the first stage is reported below the table.

Table 2.6: Adding Endogenous Expenditure Control (OLS)

	OLS Regressions			
	FHFA	Zillow	FHFA	Zillow
	(1)	(2)	(3)	(4)
Delta(UAAL per capita), 2007-2010	−0.0006	−0.0013		
	(0.0021)	(0.0023)		
Delta(UAAL per capita), 2007-2009			−0.0030	−0.0039
			(0.0032)	(0.0033)
Per Capita Expenditures, 2010	0.7415**	0.7011**	0.9473***	0.9077***
	(0.2979)	(0.3253)	(0.3198)	(0.3385)
Constant	13.4062***	22.3456***	13.0494***	21.9844***
	(1.9959)	(2.3477)	(1.9571)	(2.2973)
Observations	105	105	105	105
R^2	0.0717	0.0352	0.0792	0.0426
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table augments the baseline specification of Table 2.3 with one additional regressor: per-capita county-level public expenditures on current goods and services. This measure is computed from Census of Governments Government Finance Statistics by aggregating "Direct Expenditures" over all governments within a county and then normalizing by county population. All estimates are multiplied by 100. All models are estimated using OLS. The dependent variable is the log growth of a county-level home price index (HPI). Columns (1) and (3) use the Federal Housing Finance Authority HPI for the dependent variable, and columns (2) and (4) use Zillow's ZHVI HPI. Columns (1) and (2) present results using changes in per-capita UAAL between 2007 and 2010. Columns (3) and (4) present results using the changes in UAALs between 2007 and 2009. Heteroskedasticity-robust standard errors are reported for all specifications.

Table 2.7: Adding Endogenous Expenditure Control (IV)

	2SLS Regressions			
	FHFA	Zillow	FHFA	Zillow
	(1)	(2)	(3)	(4)
Delta(UAAL per capita), 2007-2010	−0.0047	−0.0075		
	(0.0065)	(0.0083)		
Delta(UAAL per capita), 2007-2009			−0.0096	−0.0154
			(0.0141)	(0.0191)
Per Capita Expenditures, 2010	1.2346	1.4632	1.5590	1.9858
	(0.7982)	(1.0491)	(1.3270)	(1.8369)
Constant	12.5784***	21.0659***	11.9982***	20.1315***
	(2.3312)	(2.9061)	(3.0203)	(3.9851)
1st Stage Partial F-Stat	5.00	5.00	1.83	1.83
Observations	105	105	105	105
R^2	0.0408	−0.0159	0.0403	−0.0413

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table shows the IV counterpart to the OLS estimates in Table 2.6. All estimates are multiplied by 100. All specifications are estimated using 2SLS. The dependent variable is the log growth of a county-level home price index (HPI). Columns (1) and (3) use the Federal Housing Finance Authority HPI for the dependent variable, and columns (2) and (4) use Zillow's ZHVI HPI. Columns (1) and (2) present results using changes in per-capita UAAL between 2007 and 2010. Columns (3) and (4) present results using the changes in UAALs between 2007 and 2009. Per-capita expenditures are computed from Census of Governments Government Finance Statistics by aggregating "Direct Expenditures" over all governments within a county and then normalizing by county population. Heteroskedasticity-robust standard errors are reported for all specifications. The partial F-statistic from the first stage is reported below the table, and is also computed using robust standard errors.

Table 2.8: Contemporaneous Changes in Home Prices and UAALs

	Dependent variable:			
	2008-2009 Growth in FHFA HPI			
	OLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)
Delta(UAAL per capita), 2007-2009	0.001	0.0005	0.0002	−0.004
	(0.001)	(0.001)		
Per Capita Expenditures, 2009		0.071		0.476
		(0.148)		(0.462)
Constant	-9.848^{***}	-10.105^{***}	-9.528^{***}	-10.849^{***}
	(0.928)	(0.936)	(1.030)	(1.372)
1st Stage Partial F-Stat	n/a	n/a	5.81	1.97
Observations	105	105	105	105
R^2	0.007	0.008	0.003	−0.047
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results from regressing contemporaneous changes in house prices on changes in unfunded pension liabilities. All estimates are multiplied by 100. Columns (1) and (2) show results from an OLS regression. Columns (3) and (4) present results from a 2SLS regression. The dependent variable is log changes in the FHFA county-level HPI during 2008 to 2009. The independent variable is changes in UAAL from (end of year) 2007 to 2009. The specifications in column (2) and (4) also control for total county per-capita expenditures in 2009. Heteroskedasticity-robust standard errors are reported for all specifications. The partial F-statistic from the IV first stage is reported in columns (3) and (4).

Table 2.9: Effect of UAALs on Home Prices Using Individual Properties (OLS)

	OLS Regressions			
	Growth in Home Price, 2010-2016			
	(1)	(2)	(3)	(4)
Change in UAAL, 2007-2010	-0.006473^{*}	-0.007855^{*}	-0.006294^{*}	-0.007511^{*}
	(0.003534)	(0.004069)	(0.003251)	(0.003836)
Total Expenditures, 2010		10.684580^{***}		10.874260^{***}
		(2.654495)		(2.546048)
Tract: Delta Median Income			9.120298^{**}	7.443189^{*}
			(3.892944)	(4.062925)
Tract: Delta Owner Percentage			-18.310550	-18.137770
			(12.306440)	(12.339960)
Tract: Delta Unemployment			-47.783250^{**}	-53.662960^{***}
			(20.196170)	(19.031240)
Tract: Delta SNAP Share			-49.567130^{**}	-45.527910^{**}
			(20.312470)	(19.207650)
Property: Delta Square Footage			0.008508^{**}	0.008750^{**}
			(0.004037)	(0.004063)
Property: Delta Pool			1.110385	1.090318
			(3.231704)	(3.117233)
Property: Delta Fireplace			-5.421008	-4.270433
			(3.995097)	(3.856844)
Clusters	48	48	48	48
Observations	15,023	15,023	14,547	14,547
R^2	0.121640	0.130361	0.132280	0.141397
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results from OLS regressions of changes in property-level house prices between 2010 and 2016 on changes in unfunded pension liabilities between 2007 and 2010. All estimates are multiplied by 100. The unit of observation is a property for which a market transaction was observed in both 2010 and 2016. Columns (2) and (4) include a control for county-level expenditures in 2010. All tract-level variables and property-level variables are changes between 2010 and 2016. Tract level variables are continuous, as is "Delta Square Footage". "Delta Pool" and "Delta Fireplace" are 1 if a pool or fireplace was added, and -1 if a pool or fireplace was removed. Standard errors are clustered at the county level.

Table 2.10: Effect of UAALs on Home Prices Using Individual Properties (2SLS)

	2SLS Regressions			
	Growth in Home Price, 2010-2016			
	(1)	(2)	(3)	(4)
Change in UAAL, 2007-2010	0.002456	−0.003073	0.002114	−0.003368
	(0.009110)	(0.007340)	(0.009034)	(0.007233)
Total Expenditures, 2010		10.359000***		10.668730***
		(2.480936)		(2.368129)
Tract: Delta Median Income			8.851393**	7.046032*
			(4.069037)	(4.184771)
Tract: Delta Owner Percentage			−17.824190	−18.983190
			(12.015690)	(11.756340)
Tract: Delta Unemployment			−47.988360**	−53.653890***
			(19.994220)	(18.880030)
Tract: Delta SNAP Share			−46.716870**	−43.101620**
			(20.132820)	(18.994980)
Property: Delta Square Footage			0.008589**	0.008799**
			(0.004040)	(0.004070)
Property: Delta Pool			1.241259	1.095707
			(3.279619)	(3.126758)
Property: Delta Fireplace			−5.789999	−4.440383
			(4.064885)	(3.856190)
Clusters	48	48	48	48
1st Stage F-Stat	6.28	5.66	6.35	5.71
Observations	15,023	15,023	14,547	14,547
R^2	0.118404	0.129442	0.129748	0.141092
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results from an IV regression of changes in property-level house prices between 2010 and 2016 on changes in unfunded pension liabilities between 2007 and 2010. All estimates are multiplied by 100. The unit of observation is a property for which a market transaction was observed in both 2010 and 2016. Columns (2) and (4) include a control for county-level expenditures in 2010. All tract-level variables and property-level variables are changes between 2010 and 2016. Tract level variables are continuous, as is "Delta Square Footage". "Delta Pool" and "Delta Fireplace" are 1 if a pool or fireplace was added, and -1 if a pool or fireplace was removed. All specifications are estimated with 2SLS. The partial F-statistic from the first stage regression is reported under the table. Standard errors are clustered at the county level.

Table 2.11: Effect of UAALs on Home Prices by Home Value (OLS)

	OLS Regressions			
	Growth in Home Price, 2010-2016			
	(1)	(2)	(3)	(4)
Delta(UAAL) High Value	-0.020251^{***}	-0.021676^{***}	-0.019917^{***}	-0.021121^{***}
	(0.004318)	(0.005334)	(0.004090)	(0.005157)
Delta(UAAL) Low Value	0.006997	0.005639	0.006617	0.005392
	(0.005144)	(0.004935)	(0.004823)	(0.004666)
Total Expenditures, 2010		10.756240^{***}		10.852990^{***}
		(2.646661)		(2.582867)
Tract: Delta Median Income			11.331690^{***}	9.656104^{**}
			(3.781378)	(3.972394)
Tract: Delta Owner Percentage			-12.600070	-12.432180
			(11.928980)	(11.927570)
Tract: Delta Unemployment			-32.200790^{*}	-38.081420^{**}
			(19.409630)	(18.206740)
Tract: Delta SNAP Share			-52.251540^{***}	-48.218090^{**}
			(20.199940)	(19.067980)
Property: Delta Square Footage			0.008629^{**}	0.008870^{**}
			(0.004054)	(0.004085)
Property: Delta Pool			2.823456	2.802063
			(3.444140)	(3.312076)
Property: Delta Fireplace			-3.163871	-2.017346
			(3.847990)	(3.695701)
Clusters	48	48	48	48
Observations	15,023	15,023	14,547	14,547
R^2	0.145475	0.154314	0.154635	0.163716

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: This table repeats the regressions of Table 2.9 with the single change of partitioning the sample of properties into high-value or low-value. Properties are classified relative to the within-county sample median of 2010 home value. All estimates are multiplied by 100, and all specifications are estimated with OLS. Standard errors are clustered at the county level.

Table 2.12: Effect of UAALs on Home Prices by Home Value (2SLS)

	2SLS Regressions			
	Growth in Home Price, 2010-2016			
	(1)	(2)	(3)	(4)
Delta(UAAL) High Value	−0.012173	−0.017745*	−0.011724	−0.017171*
	(0.010626)	(0.009357)	(0.010443)	(0.009100)
Delta(UAAL) Low Value	0.017002*	0.011440	0.015960	0.010554
	(0.010326)	(0.008536)	(0.010203)	(0.008504)
Total Expenditures, 2010		10.429050***		10.585070***
		(2.477703)		(2.414813)
Tract: Delta Median Income			11.670160***	9.923922**
			(3.959626)	(4.061834)
Tract: Delta Owner Percentage			−9.940173	−10.927900
			(12.864930)	(12.637990)
Tract: Delta Unemployment			−31.737220*	−37.334320**
			(18.423600)	(17.338170)
Tract: Delta SNAP Share			−51.339110**	−47.908140**
			(20.655290)	(19.486210)
Property: Delta Square Footage			0.008692**	0.008899**
			(0.004056)	(0.004089)
Property: Delta Pool			3.138615	3.005407
			(3.516744)	(3.366877)
Property: Delta Fireplace			−3.509074	−2.171728
			(3.877983)	(3.675336)
Clusters	48	48	48	48
Observations	15,023	15,023	14,547	14,547
R^2	0.142031	0.153245	0.151471	0.162829
Note:				*p<0.1; **p<0.05; ***p<0.01

Note: This table repeats the regressions of Table 2.10 with the single change of partitioning the sample of properties into high-value or low-value. Properties are classified relative to the within-county sample median of 2010 home value. All estimates are multiplied by 100, and all specifications are estimated with 2SLS. Standard errors are clustered at the county level.

Bibliography

Ananat, E. O. (2011). The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2):34–66.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arnold, D., Dobbie, W., and Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics*, 133(4):1885–1932.

Atuahene, B. (2017). Our taxes are too damn high: Institutional racism, property tax assessments, and the Fair Housing Act. *Nw. UL Rev.*, 112:1501.

Atuahene, B. and Berry, C. (2019). Taxed out: Illegal property tax assessments and the epidemic of tax foreclosures in Detroit. *UC Irvine Law Review*, 9(4):847.

Baar, K. K. (1981). Property tax assessment discrimination against low-income neighborhoods. *The Urban Lawyer*, pages 333–406.

Barro, R. J. (1974). Are government bonds net wealth? *Journal of political economy*, 82(6):1095–1117.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2018). Consumer-lending discrimination in the era of FinTech. *Unpublished working paper. University of California, Berkeley.*

Bayer, P., Casey, M., Ferreira, F., and McMillan, R. (2017). Racial and ethnic price differentials in the housing market. *Journal of Urban Economics*, 102:91–105.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of political economy*, 115(4):588–638.

Bayer, P. and McMillan, R. (2005). Racial sorting and neighborhood quality. Technical report, National Bureau of Economic Research.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Black, D. E. (1977). Property tax incidence: The excise-tax effect and assessment practices. *National Tax Journal*, pages 429–434.

Borusyak, K., Hull, P., and Jaravel, X. (2018). Quasi-experimental shift-share research designs. Technical report, National Bureau of Economic Research.

Boyer, C. (2020). Public pensions and state government borrowing costs. *University of Chicago working paper*.

Brown, J. R. and Wilcox, D. W. (2009). Discounting state and local pension liabilities. *American Economic Review*, 99(2):538–42.

Card, D. and Lemieux, T. (1996). Wage dispersion, returns to skill, and black-white wage differentials. *Journal of Econometrics*, 74(2):319–361.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics*, 123(1):177–218.

Card, D. and Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics*, 91(11-12):2158–2184.

Cellini, S. R., Ferreira, F., and Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1):215–261.

Charles, K. K. and Guryan, J. (2008). Prejudice and wages: An empirical assessment of Becker's The Economics of Discrimination. *Journal of political economy*, 116(5):773–809.

Charles, K. K. and Hurst, E. (2002). The transition to home ownership and the black-white wealth gap. *Review of Economics and Statistics*, 84(2):281–297.

Chetty, R., Hendren, N., Jones, M. R., and Porter, S. R. (2018). Race and economic opportunity in the United States: An intergenerational perspective. Technical report, National Bureau of Economic Research.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (1999). The rise and decline of the American ghetto. *Journal of political economy*, 107(3):455–506.

Doerner, W. M. and Ihlanfeldt, K. R. (2014). An empirical analysis of the property tax appeals process. *Journal of Property Tax Assessment & Administration*, 11(4):5–34.

Donohue III, J. J. and Heckman, J. (1991). Continuous versus episodic change: The impact of civil rights policy on the economic status of blacks. Technical report, National Bureau of Economic Research.

Engle, R. F. (1975). De facto discrimination in residential assessments: Boston. *National Tax Journal*, pages 445–451.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of*.

Froot, K. A. and Frankel, J. A. (1989). Forward discount bias: Is it an exchange risk premium? *The Quarterly Journal of Economics*, 104(1):139–161.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., and Walther, A. (2018). Predictably unequal? The effects of machine learning on credit markets. *The Effects of Machine Learning on Credit Markets (November 6, 2018).*

Gittleman, M. and Wolff, E. N. (2004). Racial differences in patterns of wealth accumulation. *Journal of Human Resources*, 39(1):193–227.

Gyourko, J. and Tracy, J. (1991). The structure of local public finance and the quality of life. *Journal of Political Economy*, 99(4):774–806.

Howard, T. and Morse, A. (2020). Luxury or necessity: Public goods and services curtailments in the wake of COVID-19. *UC Berkeley Haas School of Business working paper.*

IAOO (2018). Standard on automated valuation models. *International Association of Assessing Officers.*

Kahrl, A. W. (2016). The power to destroy: Discriminatory property assessments and the struggle for tax justice in Mississippi. *Journal of Southern History*, 82(3):579–616.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10.

Knowles, J., Persico, N., and Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, 109(1):203–229.

Lichter, D. T., Parisi, D., Grice, S. M., and Taquino, M. C. (2007). National estimates of racial segregation in rural and small-town America. *Demography*, 44(3):563–581.

Logan, T. D. and Parman, J. M. (2017). The national rise in residential segregation. *The Journal of Economic History*, 77(1):127–170.

Lucas, D. and Zeldes, S. P. (2006). Valuing and hedging defined benefit pension obligations—the role of stocks revisited. *Northwestern University and Columbia University, working paper, September.*

Massey, D. S. and Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass.* Harvard University Press.

McMillen, D. P. and Weber, R. N. (2008). Thin markets and property tax inequities: A multinomial logit approach. *National Tax Journal*, pages 653–671.

Munnell, A. H., Haverstick, K., and Soto, M. (2007). Why have defined benefit plans survived in the public sector.

Myers, S. (2019). Public employee pensions and municipal insolvency. *Stanford University working paper.*

Novy-Marx, R. and Rauh, J. (2011). Public pension promises: how big are they and what are they worth? *The Journal of Finance*, 66(4):1211–1249.

Novy-Marx, R. and Rauh, J. D. (2009). The liabilities and risks of state-sponsored pension plans. *Journal of Economic Perspectives*, 23(4):191–210.

Nunnally, S. C. (2012). *Trust in Black America: Race, Discrimination, and Politics*. NYU Press.

Paglin, M. and Fogarty, M. (1972). Equity and the property tax: A new conceptual focus. *National Tax Journal*, pages 557–565.

Rauh, J. D. (2016). Hidden debt, hidden deficits. *Hoover Institution, April*.

Ricardo, D. (1824). *Funding system*.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 90(6):1257–1278.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55.

Ross, R. (2017). The impact of property tax appeals on vertical equity in Cook County, IL. *Univerity of Chicago, Harris School of Public Policy Working Paper*.

Rothstein, R. (2017). *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright Publishing.

Schulman, K. A., Berlin, J. A., Harless, W., Kerner, J. F., Sistrunk, S., Gersh, B. J., Dube, R., Taleghani, C. K., Burke, J. E., Williams, S., et al. (1999). The effect of race and sex on physicians' recommendations for cardiac catheterization. *New England Journal of Medicine*, 340(8):618–626.

Staiger, D. and Stock, J. H. (1994). Instrumental variables regression with weak instruments. Technical report, National Bureau of Economic Research.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics*, 118:26–40.

Stevens, N. (2006). Changing postal zip code boundaries. *Congressional Research Services Report for Congress*.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear IV regression. Technical report, National Bureau of Economic Research.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, 64(5):416–424.

Weber, R. N. and McMillen, D. P. (2010). Ask and ye shall receive? predicting the successful appeal of property tax assessments. *Public Finance Review*, 38(1):74–101.

Williams, D. R. (2012). Miles to go before we sleep: Racial inequities in health. *Journal of Health and Social Behavior*, 53(3):279–295.

Appendices

Appendix A

The Assessment Gap: Racial Inequalities in Property Taxation

A.1 Appendix Figures and Tables

Table A1: Assessment Ratio Differentials in California

	Assessment Value / Market Value		
	(1)	(2)	(3)
Black Mortgage Holder	0.0413***		
	(0.0101)		
Black or Hispanic Mortgage Holder		0.1060***	
		(0.0044)	
Other Nonwhite Mortgage Holder			0.0653***
			(0.0030)
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	5603	5603	5603
Observations	1,186,388	1,186,388	1,186,388
R^2	0.3816	0.3820	0.3820
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results of our baseline assessment gap analysis for California alone. We regress the log assessment ratio on a jurisdiction-year fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect an assessment ratio differential for the given grouping of minority residents relative to non-Hispanic white residents. Standard errors are clustered at the jurisdiction level.

Table A2: Assessment Gap, Using Counties instead of Taxing Jurisdictions

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Mortgage Holder	0.1687***		
	(0.0187)		
Black or Hispanic Mortgage Holder		0.1356***	
		(0.0138)	
Other Nonwhite Mortgage Holder			0.0321***
			(0.0024)
Fixed Effects	County-Year	County-Year	County-Year
No. Clusters	1982	1982	1982
Observations	6,987,915	6,987,915	6,987,915
R^2	0.8507	0.8508	0.8508
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table repeats our baseline assessment gap analysis, but uses county-year fixed effects rather than jurisdiction-year. We regress the log assessment ratio on a county-year fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect an assessment ratio differential for the given grouping of minority residents, relative to non-Hispanic white residents. Standard errors are clustered at the county level. This specification shows that our results are not driven by the way we form jurisdictions. Our preferred specifications all use the more rigorous within-jurisdiction analysis.

Table A3: Robustness: Jurisdiction-Month-Year Fixed Effects

	log(Assessment) - log(Market)		
	(1)	(2)	(3)
Black Mortgage Holder	0.1283***		
	(0.0174)		
Black or Hispanic Mortgage Holder		0.0988***	
		(0.0124)	
Other Nonwhite Mortgage Holder			0.0282***
			(0.0019)
Fixed Effects		Jurisdiction-Month-Year	
No. Clusters	37723	37723	37723
Observations	6,987,915	6,987,915	6,987,915
R^2	0.9000	0.8999	0.8999
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table repeats our baseline assessment gap analysis, but uses jurisdiction-month-year fixed effects instead of jurisdiction-year fixed effects. We regress the log assessment ratio on a jurisdiction-month-year fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table reflect an assessment ratio differential for the given grouping of minority residents relative to non-Hispanic white residents. Standard errors are clustered at the jurisdiction level. This specification shows that measurement error introduced by forming fixed effects with calendar years rather than (unobserved) fiscal years does not lead to meaningfully different estimates.

Table A4: Assessment Gap Using Simple Ratios

	Assessment Value / Market Value		
	(1)	(2)	(3)
Black Mortgage Holder	0.0897***		
	(0.0057)		
Black or Hispanic Mortgage Holder		0.0696***	
		(0.0039)	
Other Nonwhite Mortgage Holder			0.0208***
			(0.0010)
Fixed Effects	Jurisd-Year	Jurisd-Year	Jurisd-Year
No. Clusters	37723	37723	37723
Observations	6,987,915	6,987,915	6,987,915
R^2	0.6987	0.6986	0.6986
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the results of our baseline assessment gap analysis using simple ratios (assessment divided by market) as the dependent variable instead of logged assessment ratios. We regress the simple assessment ratio on a jurisdiction-year fixed effect and on categorical groupings by racial and ethnic identity. In all columns, the reference group is non-Hispanic white residents, and for clarity coefficients for groups not being considered in a given column are not reported. The estimates in this table are intended to show that our baseline findings are not being mechanically generated by a Jensen's inequality term arising from taking the log of the assessment ratio. We use logged assessment ratios in our preferred specifications because the target assessment ratio varies widely across jurisdictions, and we wish to estimate proportional variation rather than variation in levels. Other than showing that inequalities do not disappear when using simple ratios, the estimates in this table have little intuition. The estimates in this table are a weighted average of absolute variation around jurisdiction means ranging from 7% to 100%. It is therefore natural that the results are lower than our baseline findings in Table 1.1. Standard errors are clustered at the jurisdiction level.

Table A5: Effective Tax Rate, One Year Before Sale

	Effective Tax Rate - One Year Before Sale (%)					
	Tax Bill	Before Exemptions	Tax Bill	Before Exemptions	Tax Bill	Before Exemptions
	(1)	(2)	(3)	(4)	(5)	(6)
Black Mortgage Holder	15.2528***	12.2586***				
	(2.0458)	(2.1646)				
Black or Hispanic Mortgage Holder			11.6826***	7.8133***		
			(1.4850)	(1.6357)		
Other Nonwhite Mortgage Holder					3.1404***	2.0352***
					(0.2550)	(0.2959)
Jurisd-Year FE	Y	Y	Y	Y	Y	Y
Other Controls	N	N	N	N	N	N
No. Clusters	26371	26371	26371	26371	26371	26371
Observations	3,373,164	3,373,164	3,373,164	3,373,164	3,373,164	3,373,164
R^2	0.6659	0.6315	0.6657	0.6312	0.6657	0.6312
Note:						*p<0.1; **p<0.05; ***p<0.01

Note: This table repeats our analysis in Table 1.14, but uses the tax bill from the year before sale. The denominator for computing the effective tax rate remains the observed market value. Coefficients are percentages. For each racial and ethnic grouping we present two sets of results. In odd columns, we show results using an effective rate computed using the observed tax bill and observed market value. Because the observed tax bill is potentially net of a wide range of exemptions, we also compute a before-exemption effective tax rate, by adding observed exemptions to the observed tax bill and then dividing by market value. We trim any observation above a calculated effective tax rate of 25% both before and net of exemptions. We believe this to be a conservative choice as 25% is far higher than any property tax rate of which we are aware (the national median is approximately 1.4%), and is more likely than not to be a data error. All specifications use jurisdiction-year fixed effects to hold constant the level of intended taxation. Standard errors are clustered at the jurisdiction level.

Table A6: Effective Tax Rate, One Year After Sale

	Effective Tax Rate - One Year After Sale (%)					
	Tax Bill	Before Exemptions	Tax Bill	Before Exemptions	Tax Bill	Before Exemptions
	(1)	(2)	(3)	(4)	(5)	(6)
Black Mortgage Holder	13.1055***	10.2602***				
	(1.8480)	(1.9628)				
Black or Hispanic Mortgage Holder			9.7809***	7.0178***		
			(1.3657)	(1.4751)		
Other Nonwhite Mortgage Holder					2.9336***	2.0251***
					(0.2023)	(0.2154)
Jurisd-Year FE	Y	Y	Y	Y	Y	Y
Other Controls	N	N	N	N	N	N
No. Clusters	26371	26371	26371	26371	26371	26371
Observations	3,373,164	3,373,164	3,373,164	3,373,164	3,373,164	3,373,164
R^2	0.7042	0.6703	0.7039	0.6701	0.7039	0.6701

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table repeats our analysis in Table 1.14, but uses the tax bill from the year after the sale. The denominator for computing the effective tax rate remains the observed market value. Coefficients are percentages. For each racial and ethnic grouping we present two sets of results. In odd columns, we show results using an effective rate computed using the observed tax bill and observed market value. Because the observed tax bill is potentially net of a wide range of exemptions, we also compute a before-exemption effective tax rate, by adding observed exemptions to the observed tax bill, and then dividing by market value. We trim any observation above a calculated effective tax rate of 25% both before and net of exemptions. We believe this to be a conservative choice as 25% is far higher than any property tax rate of which we are aware (the national median is approximately 1.4%), and is more likely than not to be a data error. All specifications use jurisdiction-year fixed effects to hold constant the level of intended taxation. Standard errors are clustered at the jurisdiction level.

Table A7: Synthetic Assessments, Stopping Growth in January Each Year

	log(Assessment) - log(Market)					
	Real Assessments			Synthetic Assessments		
	(1)	(2)	(3)	(4)	(5)	(6)
Black Mortgage Holder	0.144***			−0.040***		
	(0.015)			(0.003)		
Black or Hispanic Mortgage Holder		0.110***			−0.049***	
		(0.011)			(0.003)	
Other Nonwhite Mortgage Holder			0.031***			−0.007***
			(0.002)			(0.001)
Jurisd-Year FE	Y	Y	Y	Y	Y	Y
Other Controls	N	N	N	N	N	N
No. Clusters	18853	18853	18853	18853	18853	18853
Observations	2,135,943	2,135,943	2,135,943	2,135,943	2,135,943	2,135,943
R^2	0.910	0.910	0.910	0.692	0.693	0.693

Note: *p<0.1; **p<0.05; ***p<0.01

Note: This table shows an alternative implementation of our proposed approach for correcting the assessment gap. The analysis in Table 1.16 uses constructed assessments which increase with the zip-code HPI until the month of sale. In this table, we use constructed assessments which change only in January of each year. This more closely parallels the actual assessment practice of generating a single value each year. In this approach, when a sale occurs, the assessment is out of date by up to 12 months. Columns 1–3 are identical to Table 1.16 and show that the overall assessment gap looks similar in the subset of homes to which we can apply this approach. Columns 4–6 show the assessment gap using January-revised synthetic assessments. All specifications include jurisdiction-year fixed effects. Standard errors are clustered at the jurisdiction level.