Property Tax Sales, Private Capital, and Gentrification in the U.S. *

Cameron LaPoint†

Yale SOM

March 2023

Latest version here

Abstract

Local governments recover revenues from overdue tax bills by auctioning off super senior claims to homes at semi-annual tax lien or tax deed sales. I construct a new nationwide registry of local tax sales to examine how property tax delinquencies facilitate institutional real estate investment in major U.S. metro areas, and the effects of these acquisitions on neighborhood composition and housing disparities. Using detailed data on over 18,000 tax lien sales linked to owners' overdue tax payment histories, I document tax liens sell at a much larger haircut than mortgage foreclosed homes – for less than 10% of *ex ante* assessed value in the vast majority of cases. Prices of homes neighboring a tax lien sale property, on average, decline within the first two years of the sale. However, in gentrifying areas, large positive pricing spillovers emerge within three years, driven by investors' conversion of former tax lien properties into luxury housing and commercial amenities. Underrepresented minority homeowners are more likely to be displaced by tax delinquency and less likely to transact homes in areas containing recent tax sales to institutional buyers. Private capital's entry into the municipal finance ecosystem has amplified gentrification and the within-city Black-white wealth gap.

Keywords: property tax sales, institutional investors, private equity, gentrification, foreclosure options, municipal finance, distressed properties, racial wealth gap

JEL classifications: G23, G51, H71, R23, R31

*I would like to thank Robert Ainsworth, Leah Brooks, James Choi, Klaus Desmet (discussant), Rohan Ganduri (discussant), Stefano Giglio, Will Goetzmann, Paul Goldsmith-Pinkham, Maria Juul Hansen (discussant), Wojciech Kopczuk, Matthijs Korevaar, Song Ma, Amanda Ross (discussant), Peter Schott, Matt Spiegel, Wayne Xinwei Wan (discussant), audiences at Yale, Florida (Warrington), George Washington University, Penn State, Oklahoma State, the 6th IEB Workshop on Urban Economics in Barcelona, AREUEA International Conference at Trinity College Dublin, UEA North American Meeting at the World Bank, AREUEA Virtual Seminar, the AREUEA-ASSA Conference in New Orleans, MFA 2023, and four anonymous grant reviewers for helpful advice and comments. I greatly benefited from stimulating conversations with Roger Palmer on local property tax assessment and tax sale procedures. I thank Zillow for providing access to their ZTRAX database. The results and opinions are those of the author and do not reflect the position of Zillow Group. I am grateful to Barbara Esty for assistance accessing the CoreLogic data through Yale University Library. I thank staff at Preqin for answering my questions about the private equity real estate asset and deals data. This research was funded, in part, by the International Center for Finance at Yale SOM. Ongoing work on this project is supported by Russell Sage Foundation Grant #2110-34216. Finally, I thank a diligent team of RAs at Yale for providing research support on this project, including Mark Hodges, Evan Hu, Sasha Kolesnikova, Mingjun Sun, and Lu Zheng. First draft: January 31st, 2022.

†Yale School of Management. 165 Whitney Avenue, New Haven, CT 06511. Email: cameron.lapoint@yale.edu; Web: http://cameronlapoint.com

1 Introduction

Local governments recover missing revenues from overdue tax bills by auctioning off claims to homes in what is known as either a tax lien or tax deed sale (Kahrl 2017; Schmidt 2021). A series of *Washington Post* exposés published between 2013 and 2014 alleged that tax lien sales contribute to gentrification, widen the Black-white wealth gap, and allow private investors to expand their real estate portfolios by foreclosing on houses for a fraction of their true value before eventually converting them into commercial property developments.[1]

This paper tests the validity of these claims by evaluating the role of tax delinquencies in facilitating institutional real estate investment in major U.S. metro areas, and by examining the effects of these acquisitions on neighborhood composition and housing disparities. I combine cutting-edge causal inference methods in spatial research settings with a new nationwide, geocoded dataset merging tax sales to property transaction records, private equity real estate deals, and establishment databases. I then use this dataset to (i) document new stylized facts about this understudied type of distressed property sale, and (ii) quantify the local spillovers of tax sales in terms of house prices, development, business entry and exit, and income and race-based sorting.

Using unusually detailed records on over 18,000 tax lien sales from Washington, D.C. linked to owners' overdue tax payment histories, I document tax liens sell at a much larger haircut compared to mortgage foreclosed homes – for less than 10% of *ex ante* assessed value in over 80% of cases, with an underlying average tax debt of just $3,700. Prices of homes neighboring a tax sale property decline, on average, by 2% within the first two years of the sale, consistent with a large literature on the negative local pricing effects of home mortgage foreclosures (e.g. Anenberg & Kung 2014; Gerardi et al. 2015; Fisher, Lambie-Hanson, & Willen 2015).

However, in both recently gentrifying areas *positive* pricing effects of up to 9% emerge within three years, driven by investors' conversion of former tax lien properties into luxury housing and commercial amenities. The positive price externalities I document are particularly pronounced in up-and-coming areas with high initial mortgage foreclosure rates and high *ex ante* HUD foreclosure and vacancy risk scores, where tax delinquency is also more common. My findings are therefore suggestive of a blight reduction channel and complement recent evidence provided by Ganduri & Maturana (2021) on the positive local externalities of non-profit property rehabilitations.

Further, buyers of properties in areas surrounding tax sale events are 2 p.p. less likely to be Black or Hispanic, pointing to tax sales as drivers of within-city racial demographic change. This is a large effect relative to the average local underrepresented minority owner probability of 12% in the year prior to a tax sale. Interestingly, sellers in tax sale neighborhoods are 3 p.p. less likely to be underrepresented minorities, relative to a 10% baseline probability in the year before a tax sale occurs. Together, these results indicate an increased incidence of housing transactions between white

[1] Michael Sallah, Debra Cenziper, & Steven Rich, "Left with Nothing," Washington Post (September 8, 2013): https://www.washingtonpost.com/sf/investigative/2013/09/08/left-with-nothing/.

counterparties. Tax foreclosure sales thus accelerate ongoing crowd-out of non-white, lower-income individuals and amplify within-city inequality, but some incumbent minority homeowners who stay in formerly tax-distressed neighborhoods may benefit from the appreciation in property values.

For tax sales, the divergence in outcomes across neighborhood types is due to opportunistic developers strategically buying up distressed homes via underpriced foreclosure options in gentrifying areas and then catering to the amenity demands of newly arriving residents. The positive effects of tax foreclosures on prices are concentrated in the thinnest markets, indicating demand forces dominate the downward pressure on prices associated with supplying new housing units. This heterogeneity in spillover effects of tax sales by neighborhood type echoes conclusions from prior studies on the desirability of housing development. Asquith, Mast, & Reed (2021) show that introducing new buildings to an area with few existing multi-family options leads to lower rents, while Soltas (2021) argues subsidized housing is less cost-effective in more exclusive areas like Manhattan. Diamond & McQuade (2019) illustrate that subsidized affordable housing development is highly valued in low-income communities, yet in higher-income communities, white and higher-income households are willing to pay to live farther away from tax credit properties.

Isolating the effect of distressed sales events on nearby property values and neighborhood characteristics is challenging due to data limitations and myriad threats to identification. On the data side, tax sales are difficult, if not impossible, to observe in most property transaction databases. Tax-related foreclosures are also frequently lumped into the same category as mortgage foreclosures. Even if researchers can credibly pinpoint tax sales, additional assumptions are required to track changes in ownership through title changes on the same property over time. To overcome these issues, I scrape administrative records of tax sales from county tax office and treasury websites and merge them with CoreLogic Tax and CoreLogic Deeds, CoreLogic Involuntary Liens, and Zillow ZTRAX. The result is an original dataset with nationwide coverage of tax sales and property markets surrounding tax delinquency events.[2]

On the identification side, tax lien and deed sales are non-random – both because investors choose properties to bid on, and tax-delinquent households are likely to reside in neighborhoods with different economic fundamentals than non-delinquent households. I adopt three main strategies to account for very local trends in economic performance which might occlude the true causal effects of tax sales. First, is the "foreclosure wave" or ring regression introduced in Campbell, Giglio, & Pathak (2011), which is a covariate-adjusted method for comparing outcomes of properties close to vs. far from a sale event. Such regressions isolate the incremental impact of one more distressed sale on local outcomes in situations such as the wake of the Global Financial Crisis when foreclosures are geographically clustered (Towe & Lawley 2013; Gupta 2019). Second is the

[2] As of writing, I have compiled partial tax sale records spanning 15 years for the counties containing Washington, D.C., Detroit, Chicago, Los Angeles, San Diego, New York City, Philadelphia, and records covering the entire states of Florida and Connecticut. Zillow ZTRAX and CoreLogic Involuntary Liens contain flags for local tax liens and auction sales which allow me to identify tax sale events and extend the main empirical analysis beyond a few major metro areas. Administrative records provide additional information on investor behavior, since ZTRAX and CoreLogic do not report auction bids or tax debt amounts.

empirical derivatives generalization of the ring method developed by Diamond & McQuade (2019) which avoids parametric assumptions about the definition of "close" and "far" by continuously smoothing over distance relative to the tax sale event. Third, even though the preceding two methods control for very local time trends, two nearby homes may still differ in idiosyncratic ways. I identify a natural control group of neighborhoods surrounding tax delinquent properties that were redeemed just before reaching auction. This approach avoids biases due to spatial spillovers by defining treatment and control groups on a basis other than distance (Pollmann 2021).

My empirical approaches account for confounding sources of local variation in neighborhood conditions and imply that tax lien sales to institutional investors contribute to new gentrification rather simply magnifying pre-existing demographic trends. Zooming in at close distances and using a combination of empirical derivatives and semi-parametric methods to compare to far (> 0.5 miles) away properties, there is no discernible pre-trend in pricing or the racial composition of buyers and sellers, implying that investors cater to white homebuyers in tax sale neighborhoods and generate very local positive pricing spillovers rather than simply timing the market. In a placebo test where I categorize neighborhoods based on population flows from 1990-2005 – before the start of my tax sale sample – instead of 2005-2019, I find much weaker, statistically insignificant price dynamics.

But what is gentrification? This in itself is a major research topic and the source of deep debates in sociology dating back to the 1980s (e.g. Marcuse 1985, 1986). In urban economics, the focus has been on how structural models can inform the selection of neighborhood characteristics which can forecast, or in the case of Glaeser, Luca, & Moszkowski (2020), nowcast gentrification using high-frequency indicators such as Yelp reviews. Examples of key state variables predicting neighborhood change include housing prices (Guerrieri, Hartley, & Hurst 2013), building depreciation (Brueckner & Rosenthal 2009), opportunity costs of highly educated workers (Su 2022), and demand for local amenities stratified by educational attainment (Couture & Handbury 2019) and race (Baum-Snow & Hartley 2020).

I use a two-stage reduced form model based in the sociological tradition of "stage models" to classify all U.S. Census tracts over the period 1990 to 2019.[3] The model admits four types of neighborhood migration patterns: abandonment, gentrifying, growth, and low-income concentration. Threshold parameters control the stringency of each type definition and an "unclassified" category which captures neighborhoods not undergoing significant demographic shifts. The first stage of the model asks whether the local economy is expanding or contracting, and the second stage further distinguishes between whether expanding or contracting areas accommodate low-income population growth. For a wide range of parameter values, 7-9% of tracts underwent gentrification during 1990–2005, compared to 11-15% between 2005 and 2019. The model

[3]Recent examples of stage models in the sociology literature include Hochstenbach & van Gent (2015), who characterize Amsterdam neighborhoods as gentrifying or non-gentrifying based on initial income in the first stage and income growth in the second stage. Hwang & Ding (2020) use the same sorting criteria in their first stage, but add housing costs and growth in the share of college-educated residents as conditioning variables for the second stage sort. The Urban Displacement Project at UC Berkeley uses a more complex sort based on initial income, current income, home values, and population growth rates to produce its 8-type model (Thomas et al. 2020).

thus reflects the well-documented, steady rise in spatial inequality in per capita incomes since the 1980s (e.g. Ganong & Shoag 2017; Gaubert et al. 2021).

To illustrate, Figure 1 compares the footprint of tax lien sales across Washington, D.C. Census tracts before vs. after the 2008 Financial Crisis, applying definitions from my two-stage classification model. Institutional investors almost exclusively purchase liens on tax-delinquent properties in gentrifying areas, while individuals and non-profits are more likely to buy liens on properties in non-gentrifying areas. Following a spike in tax delinquency rates around the Great Recession, institutional investors became even more dominant players in the tax lien sale market, accounting for 79% of sales in 2009-2012, compared to 57% of sales in 2005-2008. These trends parallel findings that large institutional investors, or so-called "Wall Street Landlords," consolidated their positions in the housing market by buying up bank foreclosed properties after 2008 (Garriga, Gete, & Tsouderou 2020; Lambie-Hanson, Li, & Slonkosky 2022). Moreover, although forming a smaller share of tax sale buyers, non-resident individual investors also produce positive pricing spillovers in gentrifying areas, consistent with the findings of Chinco & Mayer (2016), Bayer et al. (2020), and García (2022), on the contributions of out-of-town second home investors to local booms in the 2000s.

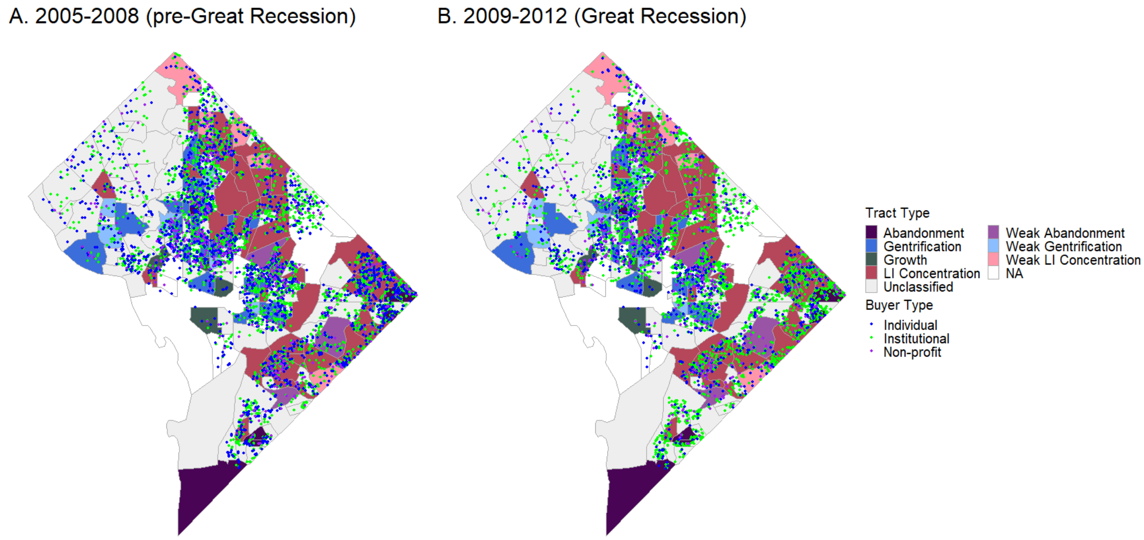
Compared to individual or non-profit investors, institutions are more likely to buy in bulk and invest in tax claims to homes located in currently gentrifying areas or neighborhoods bordering gentrifying areas. The top 15 LLCs account for 45% of the $302 million in revenue collected from tax sale auctions in D.C. These corporate retail traders at auctions act as intermediaries to larger institutional investors, often transferring liens or title to the underlying property to developers. I hand-match single-asset private equity real estate deals in Preqin to properties which show up at tax sale auctions. For the D.C. market, roughly 10% of deals, totalling $5 billion in deal value, can be matched to formerly tax-distressed multi-family and mixed-use residential property addresses.[4]

With the exception of Whitaker & Fitzpatrick (2013) and Alm et al. (2016), who focus on the Cleveland and Chicago markets, respectively, tax sales have not garnered much attention in economics. A literature at the intersection of law and public policy advocates for overhauls of tax sale systems on the grounds that the delinquency process levies unreasonably high fees on taxpayers (Rao 2012), property taxes are racially discriminatory (Hayashi 2020), and auctions erode local fiscal accountability (Baskett & Bradley 2021). The most closely related economic literature to tax sales concerns inequities in local property tax assessments. A common finding is that lower-priced homes are assessed at higher effective tax rates (Hodge et al. 2017; Berry 2018, 2021), in part due to automatic valuation models (AVMs) favored by assessors not incorporating granular geographic fixed effects (Amornsiripanitch 2021). Regressive assessments are a potential channel through which tax sales might contribute to increased marginalization in the housing market (Atuahene & Berry 2019; Ding & Hwang 2020). Wong (2020) finds that even relatively small surprise increases in

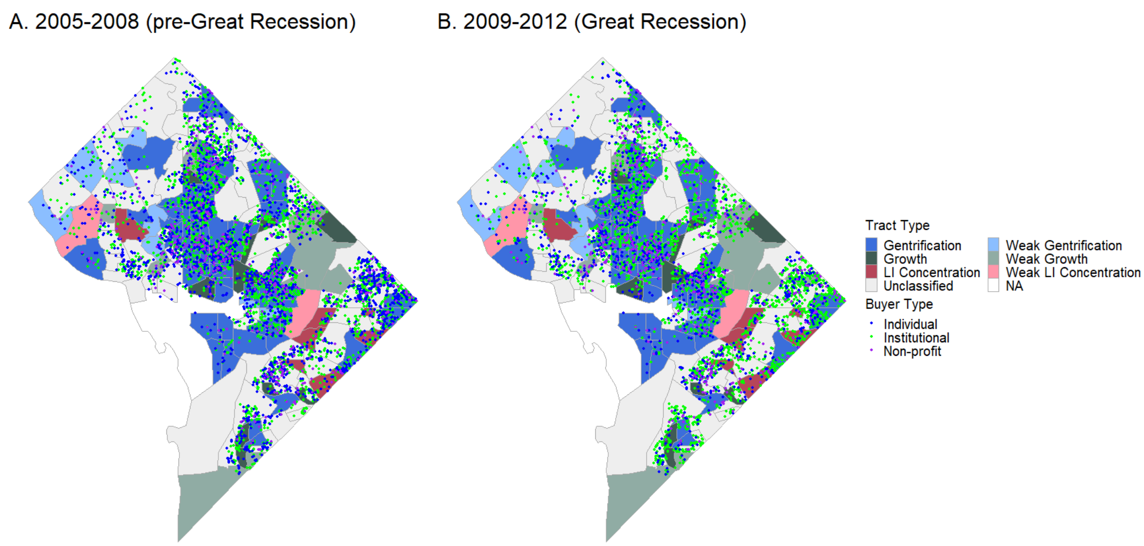
[4]According to professional investors and real estate attorneys, many private funds register for tax lien auctions under subsidiary aliases or special purpose vehicles. Anecdotal evidence suggests hedge funds may further shroud their identities by hiring college interns to register and bid on properties on behalf of the firm (Loftis 2007). These strategies to preserve corporate anonymity render it difficult to identify the ultimate owner of tax-distressed properties.

FIGURE 1. Washington, D.C. Tax Lien Sales by Census Tract Type

A. Using 1990 – 2005 Tract Classifications



B. Using 2005 – 2019 Tract Classifications



Notes: Each map overlays the location of properties tied to tax lien sales in Washington, D.C. over 2005-2008 (left) and 2009-2012 (right). Tax lien buyers classified as individuals (blue), institutions (green), or non-profits (purple) using keywords in the buyer name. Tax lien sales records obtained from D.C. Office of Tax and Revenue. I apply a generalized version of the model of income-based population sorting from Institute on Metropolitan Opportunity (2019) to classify Census tracts. See Section 3 and Appendix B text for details on the classification model. Panel A displays the tract types from estimating the model over 1990 – 2005, while Panel B shows the tract types from estimating the model over 2005 – 2019. Gentrifying tracts experience crowd-out of low-income residents, and unclassified tracts are those with little change in composition of low vs. high-income residents. I identify institutional buyers using the keywords: "LLC", "FUND", "INC", "BANK", "REALTY", "PARTNERS", "CAPITAL", "TRUST", "CORPORATION", "PLLC." I identify non-profit buyers using the keywords: "PRAYER", "CHURCH", "COMMUNITY", "FAITH", "UNIVERSITY", "COLLEGE", "SCHOOL", "BAPTIST", "FOUNDATION", "GOVERNMENT", "EMBASSY", "CENTER", "COOPERATIVE", "FRIENDSHIP", "MINISTRIES", "FEDERAL", "REHABILITATION."

monthly escrow payments translate to higher mortgage delinquency rates.

Given a dearth of prior empirical work on tax sales, I establish three key financial facts about the market for claims to tax-distressed properties. First, investors can earn returns on tax liens comparable to those for equities, even without moving to foreclose, due to a combination of high, guaranteed statutory interest rates and redemption penalties set by states to attract private capital. Second, tax and mortgage foreclosures are nearly disjoint events. Just under 1% of properties sold at tax sale experience a mortgage foreclosure during the redemption period. Legally, local tax liens are super senior and tied to the *property* rather than the individual, and thus cannot be discharged even in the event of personal bankruptcy declarations. Local governments are also required by legal due process considerations to notify lenders of possible tax foreclosure (Alexander 2000). Hence, lenders are incentivized to redeem tax debts to preserve their position in the house, so almost all severely delinquent properties are unencumbered by a mortgage. Third, in certain jurisdictions, including Washington, D.C., equity is not returned to delinquent homeowners despite the fact that annual revenue collections at auctions frequently total more than *twice* the outstanding tax debt – inclusive of interest and penalties. Local governments collect surpluses due to bidding wars between speculators, as the starting bid is usually "underquoted" relative to the appraised home value, leading to herding behavior similar to that documented by Gargano & Giacoletti (2022) for brokered home auctions, and the final bid can exceed the value of the foreclosure option.

My work adds another dimension to a burgeoning research agenda studying how the overall racial wealth gap (Derenoncourt et al. 2021) is magnified through housing market disparities, including homeownership (Perry, Rothwell, & Harshbarger 2018), housing returns (Kahn 2021; Kermani & Wong 2021), property tax assessments (Avenancio-León & Howard 2022), mortgage interest paid (Bhutta & Hizmo 2021; Bartlett et al. 2022), leverage constraints (Gupta, Hansman, & Mabille 2022), and refinancing propensities (Gerardi, Willen, & Zhang 2021; Zhang 2022). I argue that an important – but not yet probed – aspect of housing inequality is how local governments' dependence on raising property tax revenue leads to public-private partnerships benefiting certain neighborhoods and high-income in-migrants at the expense of incumbent residents, who are more likely to be low-income, mentally incapacitated elderly, and non-white taxpayers. In areas I identify as gentrifying based on incomes, the fraction of non-white homebuyers declines after a tax lien sale to an institutional buyer. These knock-on effects of local fiscal constraints on within-city and cross-racial group wealth inequality have important implications for revenue redistribution from national and state tax coffers to cash-strapped municipal treasuries.

The remainder of the paper proceeds as follows. Section 2 offers legal background on local tax administration and describes the construction of a new database of tax-distressed property sales. Section 3 proposes a two-stage model of income-stratified population flows to map gentrification across the U.S. and relates these flows to the geographic concentration of tax sales. Section 4 provides estimates for the local pricing effects of tax sales in an application to the Washington, D.C. market. Section 5 examines how the racial demographics of homeownership change around tax foreclosure events involving institutional investors. Section 6 concludes.

2 Background on Tax Lien & Tax Deed Sales

In this section, I offer background on state and local regulations governing tax sales in the U.S. I use the tax lien market in Washington, D.C. as a case study to explore potential returns accruing to investors. I then discuss how I merge newly collected data on tax sales with other property datasets to determine the effects of tax sales on home prices and neighborhood composition. Appendix A provides more details on how the statutory parameters for tax auctions and foreclosures can vary over time and across states. Given the absence of scholarly research on how tax sale regimes differ across locations, the information from this section originates from consulting the relevant state and local property tax statutes and practitioners' guides to tax lien investing (e.g. Loftis 2007).

2.1 Legal Background

By default, a lien is placed on a property once a local tax obligation is deemed delinquent. In most states, real property taxes are due on December 31, and automatically considered delinquent if unpaid by January 1. During delinquency, interest, penalties, and fees related to the administrative costs of collection accrue to the overdue bill. After a statutory delinquency period elapses, tax bills left unpaid are offered for sale through an auction.[5] Jurisdictions will then post sale notices as advertisements to investors with information on the overdue bill amounts and property location.

Legal rules for tax sales can vary at the state, city, or county level, but proceedings are largely dictated by state-level statutes. The main variation across municipalities lies in the nature of the claim being sold at auction – that is, whether the deed to the house or a lien is being sold with a foreclosure option that an investor can exercise once a required redemption period has passed. There are no legal restrictions on secondary market transfers of tax liens, but data are not systematically collected by local tax authorities on lien certificate transfers. Moreover, some states like Florida allow multiple liens to be placed on the property at a time, and the senior lienholder can exercise their foreclosure option by buying out any other stakeholders' positions after a designated holding period has passed.[6] I group tax sale methods into three major types based on legal recourse of the taxpayer and investor:

1. **Lien sales:** a lien certificate is sold at auction after a delinquency period ends. Depending on the auction rules, the lienholder may be entitled to interest and/or penalties which continue

[5]Local property tax delinquency accounts for the vast majority of local tax liens. Larger municipalities may also levy sewer, sanitation, residency, and property taxes on non-real estate assets such as cars which can result in the same tax lien and foreclosure consequences (i.e. same interest and penalty rates) if left unpaid.

[6]In Florida, local tax liens have a 7-year maturity (expiration) and 2-year redemption period before the senior lienholder can file to foreclose. Because Florida has a high maximum statutory interest rate of 18% and tax lien bids on undesirable properties go uncontested, investors face little incentive to foreclose. As of June 2021, there were over 52,000 active property liens held by investors across the 23 most populous Florida counties. Multiple liens attached to the same property are far less common in other states, in part due to lower maximum interest rates.

accruing to the property during a statutory redemption period.[7] After the redemption period ends, the lienholder can move to foreclose. One tax lien cannot wipe out another. For example, if the investor holds a local tax lien and moves to foreclose on a property which also has a federal tax lien, the IRS has 120 days to buy out the local lien at a 6% interest rate, or else the local lienholder receives the property without encumbrances. Similarly, if there are multiple local liens on the same property, then to foreclose and take possession the lienholder must first redeem the other liens.[8] For these reasons, tax liens are super senior debt claims.

2. **Deed sales:** the tax authority or other local public entity (in some cases the sheriff's office) directly forecloses on the property and then sells the deed to investors at auction. Deed sales can be conducted either in an auction setting or via "take-it-or-leave-it" offers made at over the counter sales (OTC). Localities often use OTC sales to offload property deeds which were not purchased or redeemed at previous auctions.[9]

3. **Hybrid (deed) sales:** a hybrid sale operates like a deed sale, except after the deed is transferred to a third-party investor, the former property owner or their estate can redeem the outstanding debt within a specified period to reclaim the deed. Effectively, this means the investor must wait until the redemption period ends to convert the deed claim to a title transfer. The redemption period can range from several months to years. Importantly, under the hybrid sale system, to redeem the property the previous deedholder must cover the investor's bid on the deed, rather than simply the outstanding tax debt. Thus, investors in hybrid states can create a barrier to redemption by bidding in excess of the tax debt.

Figure 2 summarizes which states fall into each auction type category. Based on the map, there are no clear divisions in preference for deed vs. lien sales based on recent state-level political leanings; 17 states plus D.C. exclusively engage in lien sales, and 23 states exclusively engage in deed sales. Florida and Kentucky are exceptions in that while they primarily operate as lien states, local governments have the authority to engage in deed sales. The remaining nine states fall into the hybrid category. Even if an investor obtains a deed through tax sale, they do not automatically hold title to the property; title companies do not accept tax deeds, and the investor would need to retain a title attorney to convert to a quit claim deed, entailing additional legal fees.

Administrative tax sale records typically list the identities of the delinquent taxpayer, the buyer, unique parcel identifiers assigned by the tax assessor's office, and in some cases additional details

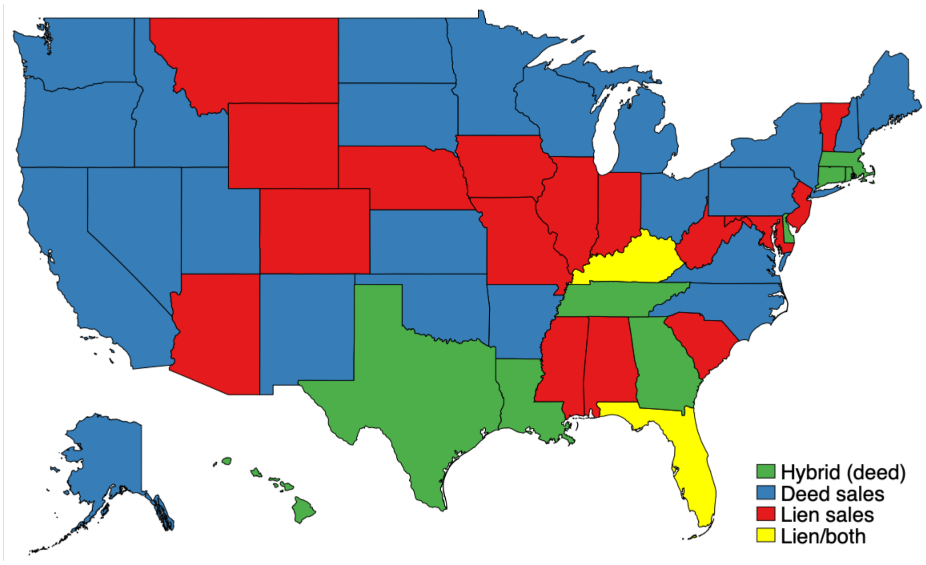
---

[7] The government is still responsible for the tax collection. If redeemed, the government then cuts a check to the lienholder. If not redeemed, then the lienholder can move to foreclose and then the property goes up for auction at what is called a scavenger or OTC sale, with the minimum bid set to the "credit bid" of tax debt + interest/penalties outstanding. Thus, the lien investor is made whole even if they do not ultimately wish to take possession of the property. Although there is variation across states in the amount of notice required, lienholders are required to file with the county to inform the taxpayer of their intent to foreclose.

[8] See IRS 5.17.2.6.5.6 Real Property Tax and Special Assessment Liens.

[9] Alabama, Arizona, Colorado, Maryland, Montana, Nebraska, South Carolina, and Wyoming sell liens at auctions but have an OTC market for initially unsold liens.

FIGURE 2. Tax Sale Auction Types by State



Notes: The map classifies states into four distinct categories based on how local governments can auction off claims to tax-delinquent properties under state law. A lien sale state is one where a lien is sold at auction after a statutory delinquency period and there is a redemption period during which the property owner is required to pay back the tax debt plus any penalties and interest. In deed sale states, the tax authority or other local entity forecloses on the property and then sells the deed to investors at auction. A hybrid state operates like a deed sale state, except after the deed is transferred to a third-party investor, the former property owner or their estate can redeem the outstanding debt within a specified period to reclaim the deed. Florida and Kentucky operate primarily as lien states but local governments have the authority to engage in deed sales. Information sourced by the author directly from the relevant statutes regarding property tax collections in each state, and from county ordinances in cases where local jurisdictions may operate in *de jure* exception to the state law. Examples of within-state exceptions include New York City and Nassau County in New York, counties in Ohio with population over 200,000 people, and counties in Pennsylvania, which can all sell liens against non-vacant property. See text and Appendix A for details on differences in how local governments may conduct auction proceedings.

such as the tax bill history attached to the property, starting bids, and any interest owed to lien investors during the redemption period. As of writing, I have compiled partial tax sale records spanning 15 years for the counties containing Washington, D.C., Detroit, Chicago, Los Angeles, San Diego, New York City, Philadelphia, and records covering the entire states of Florida and Connecticut. The remaining data collection work involves expanding the geographic and variable coverage of my tax sale database beyond the largest metros through FOIA requests and web scraping. Further, most offices maintain their raw data in non-tabular PDF format, which requires additional digitization tasks (see Appendix E for sample documents from Washington, D.C.).

Harmonizing these records into a nationwide database for research purposes involves creating a codebook accounting for other parameters besides the lien vs. deed sale type which can vary across jurisdictions and impact the composition of tax-delinquent properties observed in the data. To create a codebook, I hand-collected from property tax statutes parameters such as redemption and delinquency period lengths, interest rates on tax debt, penalty rates, and auction deposits. I provide an overview of this codebook in Appendix A by summarizing how policies governing the tax sale system influence investor returns to tax claim investing in different geographic markets.

2.2 Other Data Sources

Real estate databases. To obtain more information about the observable characteristics of the properties involved in a tax sale, including the history of title changes and mortgage lenders' claims, I link tax sale properties to CoreLogic Tax, Deeds, and Involuntary Liens by matching on both the standardized address string and the lot number. Conditioning on lot number to perform this merge resolves cases where a tax sale property address matches to multiple, yet distinct, CoreLogic properties because the tax unit of the property lies on a subdivided parcel (e.g. a multi-floor townhome consisting of separate residences).[10]

Merging to the CoreLogic suite of databases is key to my empirical strategies for identifying local spillover effects on prices and the entry and exit of homeowners by demographic group. Spatially mapping transacted properties relative to a tax sale property is important for identifying treatment and control groups and determining how the estimated effects decay with distance to the "treated" home. Unfortunately, except for in a few rare cases (e.g. Detroit), tax sale records do not include the latitude/longitude needed to create treatment and control groups based on distance. In my preliminary data collection efforts, I can match roughly 86% of tax sale properties directly to CoreLogic which contains the geo-coordinates attached to each parcel. For the remaining 14% of properties, I use Google Maps API to fill in the missing coordinates.

Each observation in CoreLogic Tax corresponds to a property assessment for tax purposes, while those in CoreLogic Deeds correspond to transactions resulting from either market-induced title changes or foreclosure events. Transactions in CoreLogic Deeds provide names of all parties involved in the transaction and detailed information about both the mortgage amount and contract type (e.g. fixed vs. adjustable rate) used to finance the purchase. In Appendix A.3, I show that it is uncommon for tax sale properties to have an attached mortgage lien. For tax sales involving a mortgage I tabulate the lender's implied equity stake in the house based on the loan amortization schedule and document that this amount is small relative to the tax debt which the lender would need to redeem to protect their claim. In my analysis of spillover effects, I restrict attention to prices for market, arms-length transactions of non-tax delinquent properties.

Merging in Involuntary Liens (IL) yields histories of tax delinquency and bankruptcy-related events attached to the property and owner(s) who face property tax distress. IL helps pinpoint

[10] In some jurisdictions, using the parcel id or full square-suffix-lot (SSL) can improve the match if this information is provided at tax sale auction, and if the parcel id attached to the tax sale follows the same conventions as the county records CoreLogic compiles. Parcel ids may not match between CoreLogic and the tax auction in cases where the authority conducting the auction is distinct from the tax authority. For instance, the sheriff's department frequently serves as the auctioneer for tax deed sales. In standardizing addresses, I conduct string operations such as removing leading zeros in the house number, and using a consistent set of street name abbreviations (e.g. "AVE" instead of "AV"). Using packages such as postmastr to parse the addresses and format them in a style consistent with how USPS formats addresses delivers a nearly identical set of merged properties.

delinquencies which did not eventually result in a lien or deed sale – something not always possible unless the tax authority maintains digitized records of tax bill mailings or payment histories. On top of the suite of CoreLogic products, Zillow ZTRAX contains auction sale flags and starting bids as well as flags for whether the property is commercial use, managed by an LLC, or if there were any recent building improvements. ZTRAX also contains a tax delinquency indicator, yet this information is subject to a "record dumping problem," whereby stated lien dates do not always match those reported in publicly available overdue notices. Hence, existing real estate databases are useful, but imperfect substitutes for the raw public records I collect in this project.

Preqin private equity deals. To help identify the ultimate owners of tax sale properties – especially those which underwent significant redevelopment – I match by hand single-asset real estate deals from Preqin to local tax sale records. I perform this merge by first linking the roster of current assets held by private equity firms to the history of deals attached to those locations according to the property name. I then match the resulting set of addresses to the tax sale rolls, which yields 43 properties (59 deals) out of 493 unique privately-held assets (696 deals) in the D.C. area over 2000-2019. This matched sample represents a small but particularly valuable fraction of all tax foreclosed properties. Deals involving a tax-distressed property account for $5 billion in deal flow. The provenance of these assets can typically not be completely tracked from the deal date back to the tax sale date; half of the matched deals feature "unidentified seller(s)." LLC investors formed to bid at tax claim auctions and developers effectively act as intermediary buyers of foreclosure options for larger institutional investors. In Appendix F, I isolate the list of 31 property deals which directly follow a tax foreclosure, totalling $3 billion in deal flow.

Business entry/exit. Since my goal is to quantify the effects of institutional capital in the tax sale property market on housing affordability and inequality, I focus primarily on homeowner demographics and prices as outcome variables. The urban economics tradition uses local amenities such as coffee shops as a way to nowcast gentrification (e.g. Glaeser, Luca, & Moszkowski 2020). I measure establishment entry and exit within tax sale neighborhoods using geocoded reviews from the Yelp public-use dataset and the Walls & Associates NETS database. Examining spillovers of tax sales to the industry composition of new businesses helps discipline the gentrification classification model based on population flows and assesses the importance of commercial development to the observed spillover effects on the housing market.

Census datasets. Finally, tract-level population flows by income from the decennial Census and 5-year ACS form the basis for the two-stage system for classifying gentrifying neighborhoods I describe in the next section. Homeowner race is not reported in standard real estate databases and in tax sale records. I assign probabilities based on owners' surnames, relying on the Bayesian approach in Imai & Khanna (2016) which links surnames to common surnames on the Census Bureau's Surname Lists and updates Bayes's rule with geocoded voter registration records whenever available.[11] Figure 3 presents sensitivity analysis from applying Bayes's rule to infer the most likely racial category for CoreLogic homeowners (Panel A) and homeowners of a property sold at tax sale (Panel B). Across a range of cutoff probabilities, homeowners with a severely delinquent property are 2 p.p. more likely to be Black and 2 p.p. less likely to be either white or Asian.[12]

2.3 Case Study: the Washington, D.C. Tax Lien Market

Besides having very detailed data which allow a decomposition of auction bids into its component parts, Washington, D.C. is representative of how tax sale auctions are conducted in the majority of states. The city uses an English price auction – sometimes called "premium bidding" in this market – and guarantees investors a statutory annual interest rate of 19.5% on tax liens. In Appendix A.2, I discuss tax auction taxonomy and provide examples of investor returns to holding lien portfolios spanning multiple property markets across states. In Appendix E I provide annotated sample tax delinquency documents which are the underlying source for the data presented in this section.

To get a sense of the revenues local governments generate from selling foreclosure options, Table 1 summarizes tax lien auction collections by year relative to total local tax revenues in that calendar year for D.C. I highlight a few key facts from the table. First, in the short-term tax auctions recover more money than the overdue tax debt inclusive of interest and penalty fees; surplus revenues as a fraction of all tax collections have stabilized around 60% in recent years, and the government regularly collects multiples of the back taxes due.[13] That is, for every one dollar of back taxes owed, the government collected 2.82 dollars in surplus revenues. Second, the surpluses originate from investors bidding in excess of the outstanding tax debt (back taxes + interest/penalties). Such excess bids are represented by the "surplus" column in the ledger. Because the starting bid is set to equal the tax debt to guarantee the government breaks even, this indicates that many auctions result in bidding wars. Although there is some variation in profit-sharing regulations across states, most statutes do not require surpluses to be automatically rebated, implying that tax delinquent

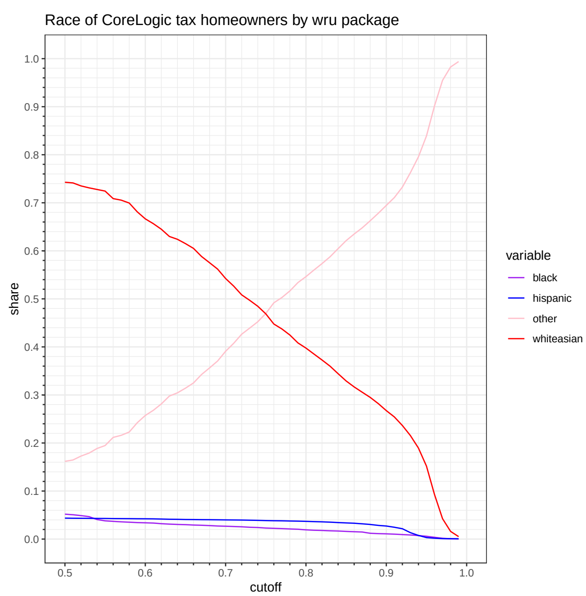
[11] Another commonly used option in the real estate literature to infer homeowner race is to merge sale records to HMDA, since mortgage lenders are required to keep track of the self-reported race of their clients. I do not pursue this route because the vast majority of tax sales do not have a mortgage attached, and given robust evidence of discrimination in mortgage approvals (e.g. Giacoletti, Heimer, & Yu 2021), racial demographics are likely to differ substantially across the two datasets.

[12] In some specifications, I use the Bayesian probabilities as outcomes or reweight the regression results instead of using biinary outcomes to avoid parametric assumptions in the assignment of race.

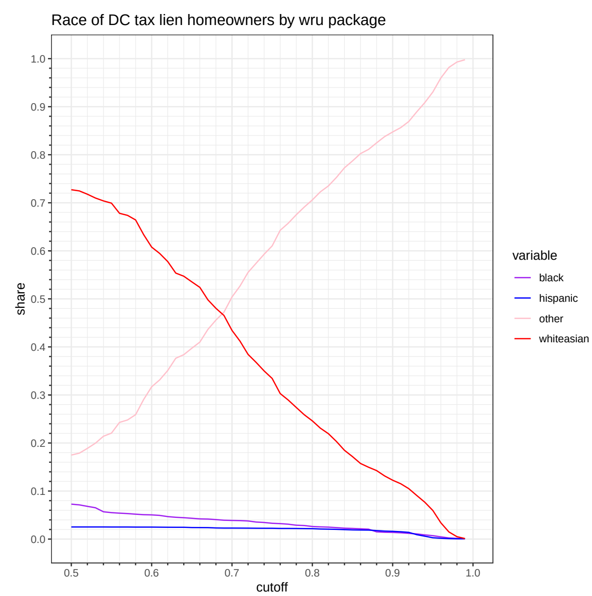
[13] This measure is left censored at zero. I calculate this net profit margin as surplus divided by total revenues, which is revenues less cost relative to revenues *for properties sold at auction.* There may be properties with a tax lien for which the government fails to sell a certificate, even at the special sales marked with asterisks in Table 1. It is not possible to systematically identify these properties because a property might show up on the initial tax sale notice but have a lien redeemed shortly before auction.

FIGURE 3. Bayesian Imputation of Homeowner Race

A. Homeowner Race in CoreLogic Tax Data



B. Homeowner Race in Tax Lien Sale Data



Notes: I use the `wru` ("Who Are You?") R package of Khanna, Imai, & Jin (2017) to assign probabilities that a homeowner is in a racial category based on matching the surname to the Census Bureau's 2010 Surname List and the Spanish Surname List. Panel A performs the imputation for all homeowners with assessed properties in the CoreLogic Tax data as of 2010, while Panel B does the same for delinquent property owners in the Washington, D.C. tax lien sale data, pooled over all years. For each owner I set the race to be the category with probability greater than the cutoff listed on the x-axis; if no such category satisfies that criterion, then I place the owner in the "other" category. The decennial Census results in a list of each surname which appears at least 100 times in the Census and the empirical probability that the respondent belongs to a self-identified racial category, including: white, Black, Asian/Native Hawaiian or Pacific Islander, American Indian and Alaska Native, multi-racial, or Hispanic/Latino. The "other" category in the figure refers to the residual proportion which are not assigned to the Black, Hispanic, or White/Asian subcategories, and therefore also includes homeowners who likely identify as multi-racial, American Indian, or Alaska Native. In cases where a homeowner has a relatively rare surname not appearing on the Surname Lists, the matching algorithm relies on race reported in voter registration records to impute probabilities. The resulting probabilities are virtually identical if I instead use the 2000 Surname Lists.

homeowners forfeit their equity (Deerson, Polk, & Martin 2023).[14]

Third, since lien sales represent instances of property owner distress, tax sale volume is countercyclical; lien volume spikes on the eve of the mortgage foreclosure crisis and remains elevated during the Great Recession. Average annual lien sale volume is roughly a third of its recession era average during the post-2012 recovery period. In Appendix D, I create a "repeat distress" index – an analog of the repeat sales index of Case & Shiller (1987, 1989) – which quantifies the value of tax claims, stripping out any variation due to time-invariant idiosyncratic characteristics of the underlying property. I show that this decline in tax sale volumes after the financial crisis coincides with a sharp decline in tax foreclosure option value to a third of its pre-recession peak, as proxied by investors' surplus or premium bid (see Appendix A.2 for a taxonomy of bidding types).

Tax liens sell at a much larger haircut than the 20-25% discount for mortgage foreclosed properties documented in the literature (e.g. Harding, Rosenblatt, & Yao 2012). The average tax lien on residences sells for \$17,400 based on an underlying \$3,700 tax debt (both in real 2012 USD). Compared to an average *ex ante* tax assessed value of \$578,100, this implies that in the absence of any legal costs, the typical investor who files for tax foreclosure can gain title for roughly 3% of the home's market value.[15] Even scaling the assessed value by the observed 11% probability of transitioning from an arms-length lien sale to a for-profit, non-bank foreclosure, results in an expected value over *three times greater* than the average auction bid.[16]

Figure 4 and Figure 5 illustrate how individual and institutional investors play different strategies at tax lien auctions, with higher returns accruing to institutional investors if they foreclose. Figure 4 compares the distribution of bid-to-value by investor type, for four different bid components: back taxes (Panel A), back taxes inclusive of interest and penalties (Panel B), bids in surplus of the tax debt (Panel C), and the overall bid which is the sum of the surplus and tax debt (Panel D). Across all four measures, there are statistically significant differences of 1-3 p.p. between average bid-to-value for individuals minus the mean for institutions. Figure 5 shows that the higher yield on a foreclosure option exercised by institutional investors is due to their targeting higher-value

[14]In Washington, D.C., the surplus is returned to investors when they file to foreclose after the redemption period or the debt is redeemed, but there is nothing in the property code providing recourse to the delinquent taxpayer (D.C. Code §§ 47-1301 – 47-1321). In contrast, in states like Connecticut, delinquent homeowners subjected to deed foreclosure must file a claim with the local superior court within 90 days, or else the profits are forfeited to the state (Conn. Gen. Stat. §§ 12-157(f),-181.). The surplus is also known as the "excess proceeds" from the tax auction.

[15]One might wonder how such a large spread between the lien bid amount and underlying asset value could prevail in equilibrium. A simple explanation is part of this spread represents an informational rent earned by savvy investors with local knowledge of neighborhoods and who specialize in distressed property claims. The D.C. OTR began posting a list of properties up for auction – along with addresses, owners' names, and tax debt amounts – starting with the July 2006 tax sale, as retrieved from a June 30, 2006 Wayback Machine archive. However, listing services like Zillow and OpenDoor, which are now widely used, only recently covered detailed characteristics and price histories for homes not listed for sale.

[16]I compute this transition probability by identifying the subset of 8,248 tax liens on residential properties which do not repeatedly show up at auction, as such repeat distress events may result from collusion between investors. I then identify 868 non-REO title transfers to non-individual entities recorded without a sales price that occur following the end of the six-month redemption period. Hence, an upper bound on for-profit investors' expected value of foreclosure relative to the bid is $(578{,}100/17{,}400) \times (868/8{,}248) \approx 3.5$.

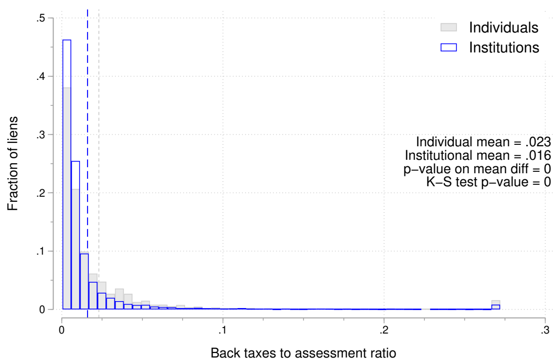
TABLE 1. Washington, D.C. Tax Sale Revenues (Millions of Dollars): 2005 – 2019

Year	# liens sold	Back taxes	Interest/penalties	Surplus	Auction revenues	Surplus-revenue ratio (%)	Total tax revenues
2005	2,181	3.52	1.04	32.76	37.35	87.71%	1,136
2006	1,997	3.81	1.07	23.06	28.77	80.16%	1,212
2007	2,083	4.31	1.65	45.11	51.82	87.06%	1,542
2008	1,366	6.05	2.27	12.46	21.31	58.46%	1,727
2009	1,068	5.91	2.11	2.33	10.61	21.99%	1,839
2010*	1,622	8.06	2.90	2.13	22.27	9.56%	1,891
2011***	1,998	6.13	2.27	4.04	13.49	29.93%	1,734
2012*	1,248	5.17	2.06	5.93	14.72	40.32%	1,880
2013	965	4.48	1.61	11.82	17.91	66.00%	1,951
2014	316	2.08	0.74	5.53	8.57	64.56%	2,035
2015*	534	2.62	1.00	8.77	12.73	68.91%	2,267
2016**	1,040	3.47	1.33	8.32	16.32	50.94%	2,364
2017*	675	2.31	0.93	9.54	13.92	68.49%	2,579
2018	516	3.51	1.21	5.89	10.66	55.23%	2,591
2019	810	5.99	3.08	12.58	21.93	57.37%	2,807
Total	18,419	67.42	25.26	190.28	302.39	62.92%	29,556

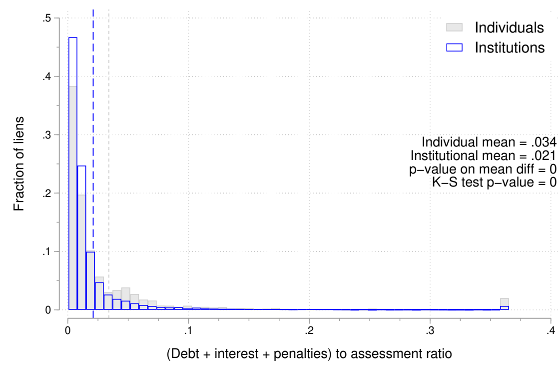
Notes: Back taxes includes real estate tax revenues *ex* interest and penalties due to the Office of Tax and Revenue (OTR). All monetary values in millions of nominal dollars. Interest/penalties includes any interest and penalties accrued on the property tax debts as of the time the property is listed on the tax sale roster. Surplus is the total revenues collected from auctions in excess of the back taxes plus interest/penalties due. Auction revenues is total revenues from tax sales, inclusive of proceeds from liens for non-property local tax bases (e.g. weed liens). Surplus-revenue ratio is surplus divided by auction revenues, which is a short-term net profit calculation implicitly treating back taxes + interest/penalties as the tax authority's operating costs. Including interest/penalties in the denominator treats these levies as administrative costs in the profit margin calculation. Finally, total tax revenues is total property tax revenues collected in that year, as reported in the Quarterly Summary of State & Local Tax Revenue (QTAX). In some years, the tax authority might convene a second or third sale on top of the regularly scheduled July auction. In such cases, I add up the subtotals for each sale event within the same year. These special sales are structured as "first come first served" in that the sale price is always a take-it-or-leave-it offer set equal to the lost revenues *ex* interest/penalty due to the treasury. Years with $*$ featured special sales, and the number of stars indicates the number of special sales. Tax auction variables sourced from the buyer's books for 2005 – 2019 available through the Washington, D.C. OTR. See Appendix E for sample tax delinquency documents accompanying a typical tax auction, including buyer's book excerpts.

FIGURE 4. Distribution of Bid-to-Value by Investor Type

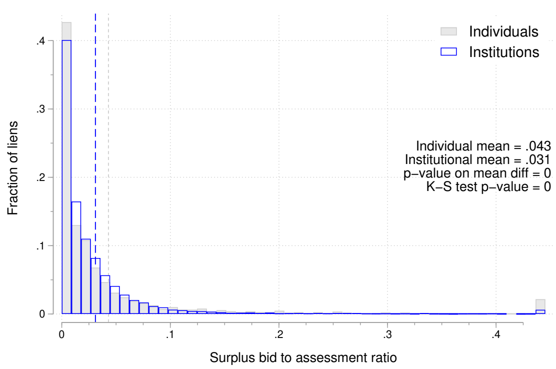
A. Back Taxes to Last Assessed Value



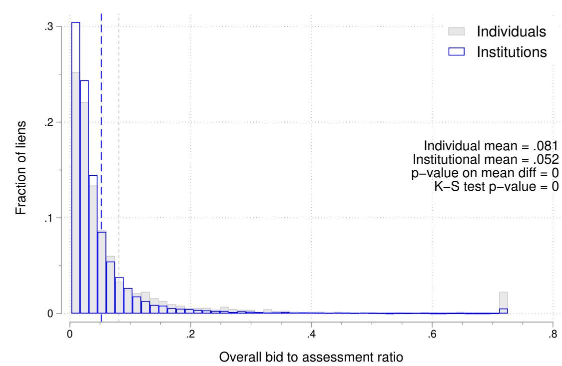
B. Tax Debt to Last Assessed Value



C. Surplus Bid to Last Assessed Value

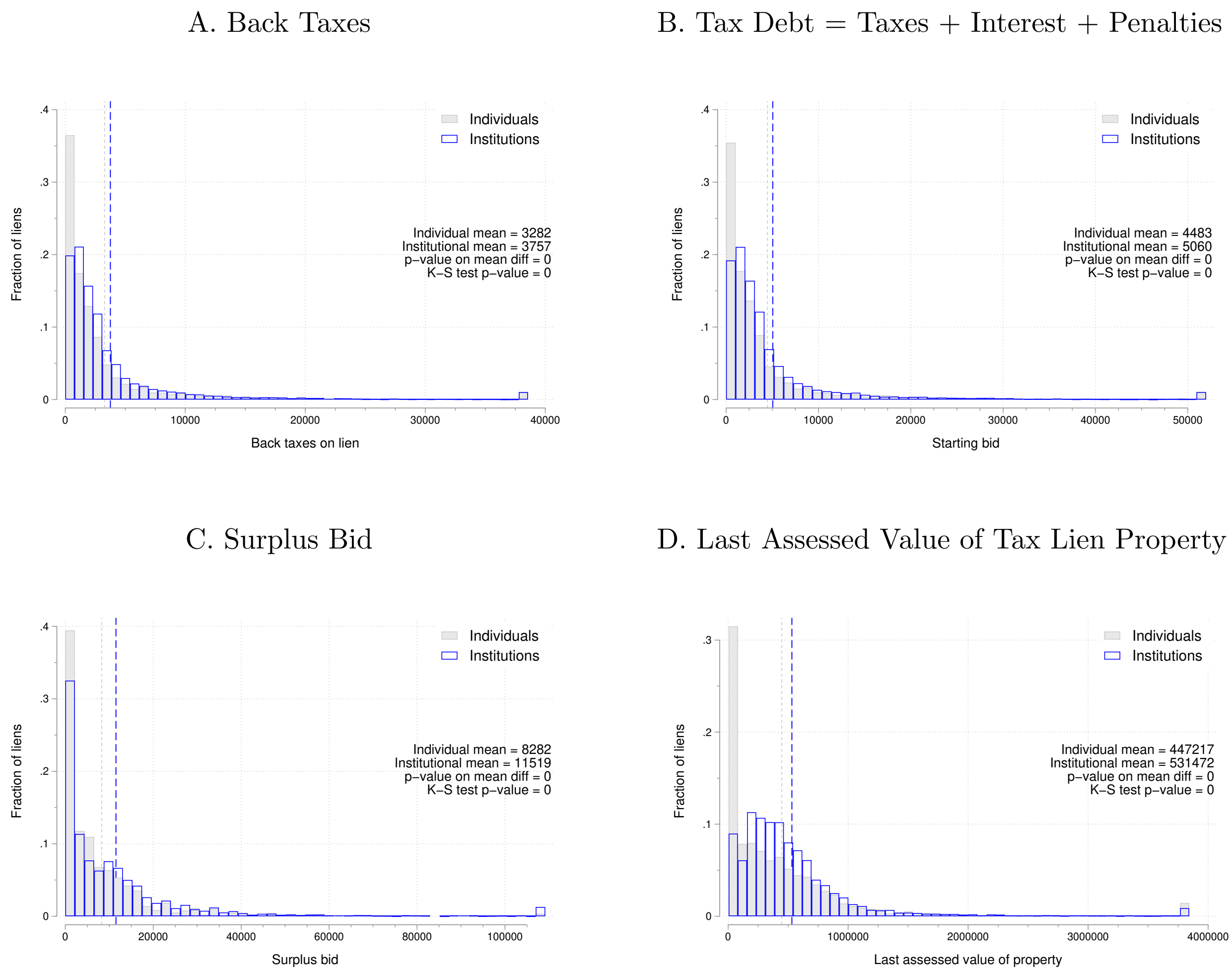


D. Overall Bid to Last Assessed Value



Notes: Each panel restricts to the tax lien sale data for Washington, D.C. and plots a measure of tax lien bid value to the assessed value for the underlying property. Panel A compares the distribution of back taxes (local property + other local taxes) to the assessed value for individual vs. institutional tax lien buyers. Panel B does the same for the total tax debt (back taxes + interest + penalties), Panel C for the surplus bid in excess of the tax debt, and Panel D for the overall bid (tax debt + surplus). In each panel, I obtain assessed values by merging each tax sale property on the lot and address combination to Zillow ZTRAX. I use the last assessed value before the tax sale, or the assessed value in year = tax sale year − 1. I identify institutional buyers using the keywords: "LLC", "FUND", "INC", "BANK", "REALTY", "PARTNERS", "CAPITAL", "TRUST", "CORPORATION", "PLLC." Individuals exclude institutional buyers and charities or other non-profit buyers identified using the keywords: "PRAYER", "CHURCH", "COMMUNITY", "FAITH", "UNIVERSITY", "COLLEGE", "SCHOOL", "BAPTIST", "FOUNDATION", "GOVERNMENT", "EMBASSY", "CENTER", "COOPERATIVE", "FRIENDSHIP", "MINISTRIES", "FEDERAL", "REHABILITATION." I winsorize each bid-to-value ratio at the 1st and 99th percentiles. I provide p-values for a test of the difference in means and Kolmogorov-Smirnov tests of the difference in the distributions between institutional and individual investors.

FIGURE 5. Distribution of Equity Stake Measures (Real 2012 Dollars) by Investor Type



Notes: Each panel restricts to the tax lien sale data for Washington, D.C. and plots a measure of investors' equity stake in the underlying property. All values in real 2012 dollars, converted from nominal terms using the PCE deflator. Panel A compares the distribution of back taxes (local property + other local taxes) for individual vs. institutional tax lien buyers. Panel B does the same for the total tax debt (back taxes + interest + penalties), Panel C for the surplus bid in excess of the tax debt, and Panel D for the last assessed property value for the property attached to the lien. In the last panel, I obtain assessed values by merging each tax sale property on the square and address combination to Zillow ZTRAX. I use the last assessed value before the tax sale, or the assessed value in year = tax sale year − 1. In Panel C, cases where surplus = 0 represent uncontested auctions, since the starting bid is set to the value of the total tax debt. I identify institutional buyers using the keywords: "LLC", "FUND", "INC", "BANK", "REALTY", "PARTNERS", "CAPITAL", "TRUST", "CORPORATION", "PLLC." Individuals exclude institutional buyers and charities or other non-profit buyers identified using the keywords: "PRAYER", "CHURCH", "COMMUNITY", "FAITH", "UNIVERSITY", "COLLEGE", "SCHOOL", "BAPTIST", "FOUNDATION", "GOVERNMENT", "EMBASSY", "CENTER", "COOPERATIVE", "FRIENDSHIP", "MINISTRIES", "FEDERAL", "REHABILITATION." I winsorize each variable at the 1st and 99th percentiles. I provide p-values for a test of the difference in means and Kolmogorov-Smirnov tests of the difference in the distributions between institutional and individual investors.

properties (Panel D) and to their higher likelihood of participating in contested auctions (Panel C). 27.9% of lien purchases by individual investors are in uncontested auctions where the total bid equals the tax debt (i.e. the starting bid), compared to only 19.1% for institutional investors. In all cases, Kolmogorov-Smirnov tests reject the null that the distributions are equivalent for individual vs. institutional investors (p-value < 0.000).

In Appendix A.3 I discuss mortgage lenders' incentives to redeem borrowers' tax debt. Using the implied amortization schedule for mortgages recorded when the delinquent owner purchased the property, I show that for over 80% of delinquent properties the loan is either paid off by the tax sale date, or otherwise it would not be profitable for the lender to redeem the tax debt. Of the fraction remaining (5% of the sample) for which I can produce a credible, strictly positive estimate of the loan balance, the median gap between the principal and tax debt is $152,000. Roughly a fifth of this already small batch of tax liens, or only 1% of the total lien sample, results in a mortgage foreclosure or short sale within the statutory six-month redemption period. Hence, mortgage foreclosure and tax foreclosure are largely disjoint events.

3 The Geography of Gentrification

This section introduces a two-stage, semi-parametric model of income-based population sorting for classifying Census tracts as gentrifying vs. non-gentrifying. The model captures long-run trends towards widening regional inequality. I use this classification model to show tax delinquency events are more likely to occur in gentrifying areas (cf. Figure 1). My methodology for identifying gentrification adapts the steps outlined in the Institute on Metropolitan Opportunity (2019) report. The basic procedure is a two-stage sort of Census tracts. In the first stage, I determine whether an area is economically expanding or contracting, while in the second stage, I further stratify expanding and contracting tracts on the basis of whether they accommodate low-income population growth.

I estimate the model over two periods: 1990 to 2005 and 2005 to 2019. I choose the 2005-2009 5-year ACS as the cutoff point because most local tax offices began recording their tax sales in a systematic way around the late 2000s. Splitting the sample into these two 15-year periods allows me to separately assess long-run demographic changes over periods prior to vs. after the tax sales I study. Such analysis is crucial to understanding whether institutional investor involvement in the tax-distressed property market exacerbated ongoing gentrification, or led to new crowd-out of low-income residents in previously affordable neighborhoods. I include all Census tracts in the country with non-missing data for each of the subperiods.[17]

[17] I acknowledge criticisms of using relatively small geographic areas such as Census tracts rather than larger areas such as zip codes as the unit of economic analysis (Small & McDermott 2006). However, income-based population variables are only publicly available in a consistent manner at the 5-digit zip code level beginning with the 2000 Census, which limits researchers' ability to explore long-run demographic changes through the lens of sorting models. Estimating the model prior to 2000 is important for determining whether institutional capital in the real estate market led to an expansion of gentrification, so I adopt the tract level as my geographic unit for classifying neighborhoods.

Before getting to the classification system, I provide some intuition for the model and situate it within two disparate literatures: a sociology literature aimed at measuring gentrification, and an economics literature which seeks to understand the sources of demographic change and urban migration patterns through a combination of empirical and structural methods. A seminal text in sociology is Marcuse (1985), who defines gentrification as a process through which

> *...new residents – who disproportionately are young, white, professional, technical, and managerial workers with higher education and income levels – replace older residents – who disproportionately are lower-income, working-class and poor, minority and ethnic group members, and elderly – from older and previously deteriorated inner-city housing in a spatially concentrated manner, that is, to a degree differing substantially from the general level of change in the community or region as a whole.*
> (Marcuse 1985, pg. 198-199)

The two-stage model I use borrows the area type names from Marcuse (1985, 1986): abandonment, gentrification, growth, and low-income concentration. Although there are many flavors to the two-stage classification approach, the basic procedures are the same; the first stage classifies geographic areas based on whether the local economy is expanding or contracting along some conditioning set of variables, while the second stage further distinguishes between whether expanding or contracting economies foster low-income population growth.[18]

As the above quoted definition highlights, this sociological concept of gentrification is very general. Urban economists have attempted to refine this definition by invoking the support of structural models which feature labor and/or firm sorting and offer state variables which can predict or nowcast gentrification (Glaeser, Kim, & Luca 2018). Many of these state variables are mentioned in the Marcusian definition, including house prices (Guerrieri, Hartley, & Hurst 2013), building age (Brueckner & Rosenthal 2009), the opportunity cost of highly educated workers (Su 2022), and race-specific demand for local amenities (Baum-Snow & Hartley 2020) like coffee shops or "fast casual" restaurants (Glaeser, Luca, & Moszkowski 2020). In Appendix B, I document concordances between the neighborhood classifications admitted by my model and these other common variables used as proxies for gentrification.

My version of the two-stage sociological model is governed by three threshold parameters and incorporates mechanisms from the canonical Rosen-Roback framework and Schelling's (1971,1978) tipping point theory of segregation. The standard optimality condition in a Rosen-Roback spatial equilibrium model dictates that individuals migrate such that their indirect utility is equal to the amenity-adjusted real wage offered by a location. While Schelling's theory refers to race-based sorting, tipping points observed in major metros along the dimension of underrepresented minority

[18]Examples of recent applications of such stage models include Hochstenbach & van Gent (2015), who stratify Amsterdam neighborhoods based on initial income and income growth, and Hwang & Ding (2020), who stratify Philadelphia neighborhoods based on initial income and then sort on growth in housing costs and growth in the share of college-educated residents.

population shares are strongly associated with tipping points in terms of income-based population shares, as evidenced in Bayer, Fang, & McMillan (2014). A version of the model in which one replaces income-based population with racial minority population variables delivers similar tract-level classifications, but has the disadvantage of hardwiring race-based sorting into the definition of gentrification. In Section 5, I show that even within income-based gentrification areas, tax sales to institutional investors result in lower probabilities of very nearby home purchases by non-white buyers.

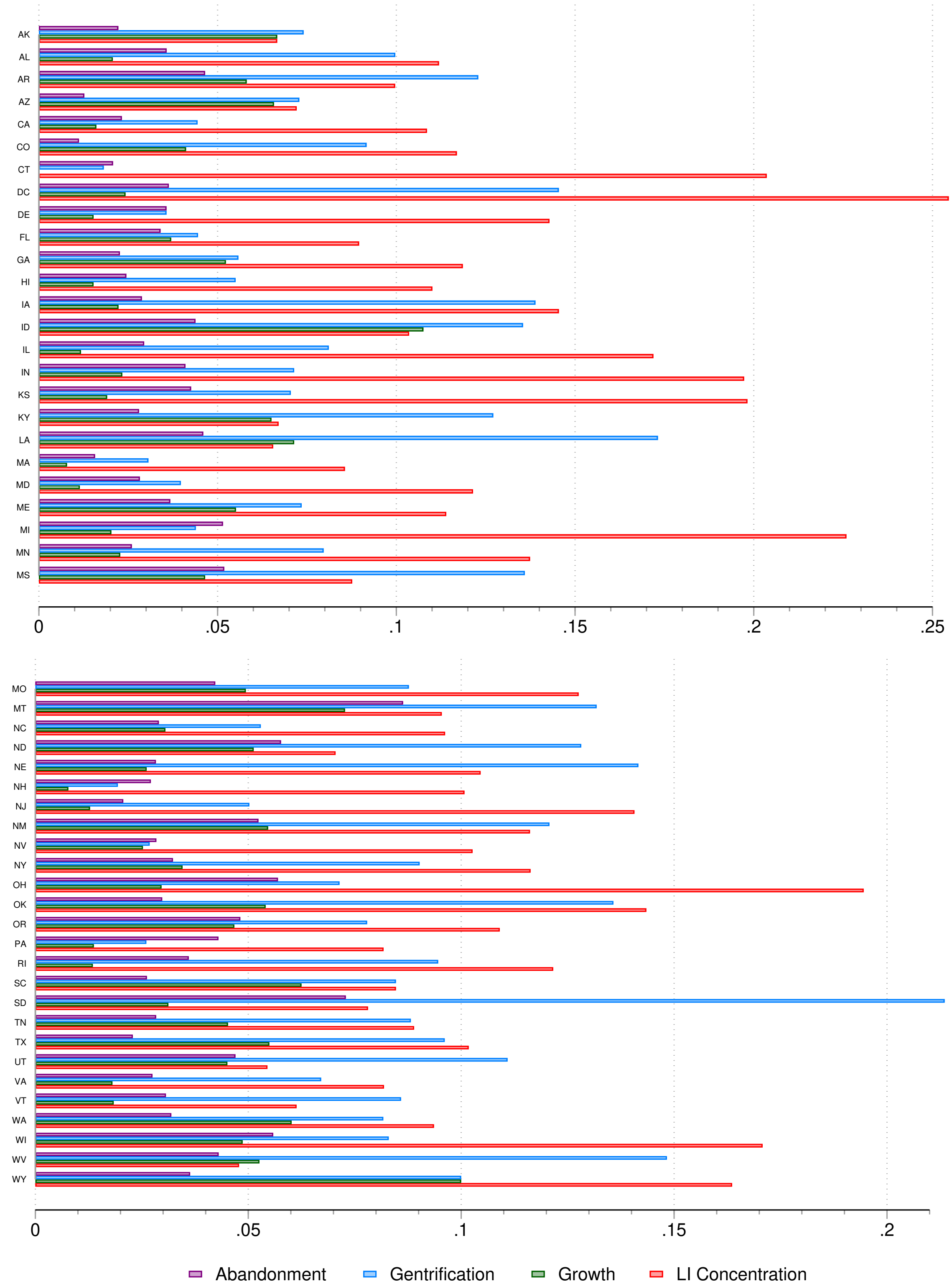
I defer the full model steps to Appendix B.1, but outline how the "tipping point" parameters $\{x_1, x_2, x_3\}$ result in a classification scheme. The first two parameters sort Census tracts on the strength of the local economy – as evidenced by in-migration patterns – while the final parameter sorts within each of those two first-stage groups on the extent of out-migration of low-income individuals. As x_1 and x_2 move towards $\pm\infty$, holding x_3 fixed, the share of neighborhoods being classified as experiencing any income composition change over the time period declines, as it becomes more difficult to attain a tipping point.

Figure 6 and Figure 7 summarize income-based population sorting in the U.S. for the baseline version of the 4-type model. As shown in Figure 6, there is substantial variation in tract type proportions across states over the 1990-2005 period. The fraction of *gentrifying* tracts ranges from 1.8% in Connecticut to 21.4% in South Dakota, compared to the national average of 7.4%. At the same time, the vast majority of tracts in the U.S. (65.6%) have stable income composition, and very few are *abandoned* (3.3%) or *growing* (3.3%) both economically and in terms of the low-income population. I present analogous results for tract type proportions by state and for the entire U.S. over the later subperiod (2005-2019) in Appendix B.3. A larger fraction of tracts have moved to *gentrification* or *LI concentration* over time. At the national level, the share of gentrifying tracts grew by 3.3 p.p. and the share of LI concentration tracts grew by 4.9 p.p. Thus the model mirrors the declining per capita income convergence and increasing regional inequality documented for the U.S. since the 1970s (Ganong & Shoag 2017; Gaubert et al. 2021).

Figure 7 indicates that there is also large variation in how tracts across different regions transition to and from gentrification. Illinois and D.C. top the list with a 27% and 26% persistence in gentrification between the two 15-year periods. Six states have negative persistence, meaning many tracts in those states have crossed tipping points since 1990. This heterogeneity in persistence upholds sociological theories which argue that gentrification is more of a short-lived phenomenon in places where newly arriving residents are risk neutral with respect to amenity deterioration (Kerstein 1990). Consequently, in relatively high amenity value urban areas such as Washington, D.C. and Chicago (Albouy 2016), gentrifiers are less likely to promote policies to accommodate low-income residents.

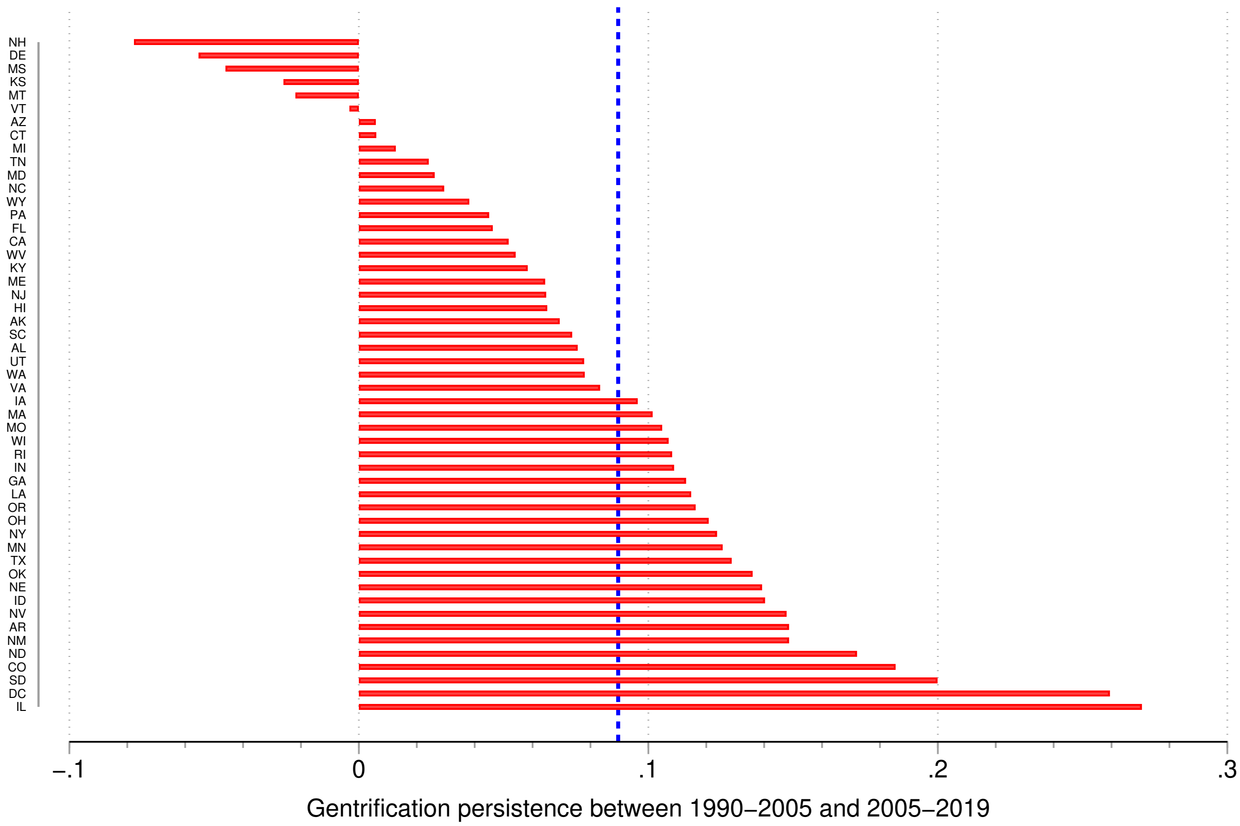
To unpack the direction of these tipping points, Table 2 presents the transition probability matrix for tract types over the entire U.S. between the two subperiods. Over one-third of tracts began with stable income sorting in the earlier period and and remained stable (*unclassified* $\longrightarrow$

FIGURE 6. Tract Type Proportions by State (1990–2005)



Notes: The figure shows the proportion of 2010 Census tracts in each state that fall into the four categories obtained from the two-stage classification model estimated over the period 1990-2005. Within each state, the tract type proportions do not sum to one, because the residual fraction of tracts fall into an *unclassified* category, which means the tract did not undergo major changes in income-based population sorting during the time period. I downloaded the 1990 decennial Census and 2005-2009 5-year ACS tract-level data from IPUMS NHGIS. See Appendix B.2 for details on the crosswalking methods.

FIGURE 7. Long-run Gentrification Persistence by State



Notes: The figure presents the within-state, cross-tract correlation between gentrification status in the 1990-2005 and 2005-2019 periods according to the two-stage classification model. In each subperiod, I create a flag equal to one if a tract is either weak or strong gentrifying, and equal to zero otherwise. The correlation of this dummy between the two subperiods indicates the joint likelihood that either a gentrifying area continues to gentrify or that a non-gentrifying area will gentrify in the future. Therefore, the measure summarizes the persistence of gentrification over several decades. The dashed blue line displays the gentrification persistence (9%) for the entire U.S. I downloaded the 1990 and 2000 decennial Census, and 2005-2009 and 2015-2019 5-year ACS tract-level data from IPUMS NHGIS. See Appendix B.2 for details on the crosswalking methods.

TABLE 2. Transition Probability Matrix: 8-type Tract Classification Model

	Abandonment	*Gentrification*	*Growth*	*LI Concentration*	*Unclassified*
Abandonment	**0.47%**	0.58%	0.15%	0.13%	2.02%
Gentrification	0.37%	**2.33%**	0.75%	1.46%	4.48%
Growth	0.17%	0.79%	**0.64%**	0.54%	2.51%
LI Concentration	0.53%	1.52%	0.033%	**6.95%**	6.45%
Unclassified	1.59%	9.69%	3.43%	14.18%	**36.76%**

Notes: The table shows the empirical probability that a 2010 Census tract transitions from one type in 1990-2005 to another type in 2005-2019. The rows indicate the initial tract type in 1990-2005, while the columns indicate the more recent tract type in 2005-2019. Each of the four categories pools the weak and strong definitions of the tract type obtained from the 8-type model. Therefore, the row totals are the fraction of tracts which satisfy the weak definition for that type in the earlier subperiod. *Unclassified* refers to a tract for which none of the 8-type tract definitions apply but for which income-based population variables are available.

unclassified). Of tracts that were initially gentrifying, a quarter remain gentrifying, while 48% reach the steady state represented by *unclassified*. Thus, for roughly half of tracts undergoing gentrification, crowd-out of low-income residents ends within a couple decades. A minority of 19% of gentrifying tracts undergo rapid reversals and become either *abandoned* or *LI concentration* tracts. Incumbent residents in gentrifying areas are, on average, unlikely to move, as also demonstrated by Ding, Hwang, & Divringi (2016) who use individual-level data on mobility from the FRBNY Consumer Credit Panel for Philadelphia.

4 Local Effects of Tax Liens on House Prices

I outline several empirical strategies for identifying the local spillover effects of tax sales. These methods include the standard parametric ring difference-in-differences (DD) which adjusts for differences in property characteristics across neighborhoods and a generalization of the ring method introduced by Diamond & McQuade (2019) which allows me to trace out how pricing externalities evolve continuously with distance from a tax sale location. Finally, I identify plausible counterfactuals for tax sale properties using properties with redeemed tax debts and machine learning techniques which address the potential for attenuated point estimates due to spillover effects between geographically defined treatment and control groups. I present results from additional ring-style difference-in-differences estimators in Appendix C.

4.1 Estimates from Foreclosure Wave Regressions

My first strategy is a covariate-adjusted version of the non-parametric ring DD based on the design in Linden & Rockoff (2008) and applied by Campbell, Giglio, & Pathak (2011) to study forced sales occurring at close proximity to each other. I estimate regressions of the following form:

$$\begin{aligned} y_{i,c,t} &= \alpha_{c,t} + \gamma_m + \beta' \cdot \mathbf{X_{i,t}} + \delta_{C,B} \cdot g(N_{C,B}; D_{C,B}) \\ &+ \delta_{C,A} \cdot g(N_{C,A}; D_{C,A}) + \delta_{F,B} \cdot h(N_{F,B}) + \delta_{F,A} \cdot h(N_{F,A}) + \varepsilon_{i,c,t} \end{aligned} \tag{4.1}$$

where $y_{i,c,t}$ is an outcome such as a dummy indicating the race of the buyer of property i, or the sale price of the property i at time t and located in Census tract c. The $\alpha_{c,t}$ represent tract $\times$ year fixed effects, and γ_m are month dummies included to remove seasonality in home sales. $\mathbf{X_{i,t}}$ is a vector of potentially time-varying controls to control for differences in property quality. $g(\cdot)$ and $h(\cdot)$ are functions which aggregate multiple tax sale events by inversely weighting sales with respect to their distance to the event property. Each function is indexed by whether i is *close* (C) or *far* (F) or occurs in a year before (B) or after (A) a proximal tax sale. A property is *close* if it is contained in the inner ring of 0.1 miles around a tax sale, but *far* if it is outside a 0.25 mile ring radius. I adopt a parameterization of the aggregator functions that renders my estimates comparable to those in the foreclosure literature. That is, $g(\cdot)$ is a distance-weighted sum where the weight is $\omega = \frac{0.1-D(i)}{0.1}$, or more generally, the inner ring radius minus the distance of each property i, divided by the inner ring radius. $h(\cdot)$ is simply an unweighted sum of all nearby tax sales.[19]

Table 3 shows results from estimating (4.1) that tax lien sales reduce market prices of closely neighboring properties by between 2.2% and 3.3%. For example, the most stringent specification in column 4 indicates that a tax sale at zero distance lowers prices of homes sold prior to the delinquency event by 2.2%; for a property 0.05 miles away from a tax sale, the estimated effect is $\frac{0.1-0.05}{0.1} \times$ 2.2% = 1.1%.[20] Even for properties 0.25 miles or more away, there are small but statistically significant negative average spillover effects ranging from 0.2% to 0.5%, as given by the difference $\delta_{F,B} - \delta_{F,A}$. Thus, in general and like mortgage foreclosures, tax delinquency is negatively capitalized in the local housing market.

However, Table 3 reveals large, positive spillovers observed in areas where tax sales are highly clustered – as captured by the differences in piecewise linear functions $\delta^p_{C,B} - \delta^p_{C,A}$ for a percentile p of the tax sale aggregator $g(\cdot)$. Tax sales are more common and nearly as geographically clustered as mortgage foreclosures or other distress events like bankruptcy or estate sales; for "close," or $g(\cdot)$ pooled across all 15 years in my sample, the 99th to 99.5th percentile is

[19] The average spillover effects captured by the difference in partial derivatives $\delta_{C,B} - \delta_{C,A}$ are largely invariant to the choice of weighting scheme, including if I apply no weights to tax sale events included in $g(\cdot)$.

[20] I provide point estimates and standard errors for all coefficients, not just the before/after differences, in Appendix C.2. Notably, the "close" and "after" coefficients are positive in the long-run, but negative in the short-run. This is suggestive evidence in favor of the redevelopment channel of tax sale foreclosure, because it means sale prices are higher for nearby properties after the tax sale event after several years have passed.

TABLE 3. Spillover Estimates of Tax Sales: Ring Difference-in-Differences

	(1)	(2)	(3)	(4)
$\delta_{C,B} - \delta_{C,A}$	-0.033^{***}	-0.032^{***}	-0.028^{***}	-0.022^{***}
	(0.010)	(0.009)	(0.006)	(0.005)
$\delta_{F,B} - \delta_{F,A}$	-0.005^{**}	-0.005^{**}	-0.002^{**}	-0.002^{**}
	(0.002)	(0.002)	(0.001)	(0.001)
$\delta^{p99}_{C,B} - \delta^{p99}_{C,A}$	0.050^{***}	0.042^{***}	0.026^{***}	0.020^{***}
	(0.006)	(0.006)	(0.004)	(0.003)
$\delta^{p99}_{F,B} - \delta^{p99}_{F,A}$	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
$\delta^{99.5}_{C,B} - \delta^{99.5}_{C,A}$	0.043^{***}	0.035^{***}	0.020^{***}	0.013^{***}
	(0.007)	(0.007)	(0.006)	(0.004)
$\delta^{99.5}_{F,B} - \delta^{99.5}_{F,A}$	-0.000	-0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)
$\delta^{99.9}_{C,B} - \delta^{99.9}_{C,A}$	0.025^{***}	0.019^{**}	0.013^{**}	0.006
	(0.009)	(0.009)	(0.006)	(0.005)
$\delta^{99.9}_{F,B} - \delta^{99.9}_{F,A}$	0.000	-0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)
Tract × year FEs	✓	✓	✓	✓
Month FEs	✓	✓	✓	✓
Multi-family dummy		✓	✓	✓
Property controls			✓	✓
Winsorized				✓
N	100,504	100,504	66,461	66,461
Adj. R^2	0.427	0.452	0.708	0.771

Notes: The table reports estimates of tax sale wave regression equation (4.1) using log property price as the outcome, and restricting to arms-length market transactions between individual homebuyers. Each specification allows for for piecewise linear effects of $g(\cdot)$ and $h(\cdot)$ over the intervals 0–99th, 99th–99.5th, and 99.5th–99.9th percentile. The reported coefficients are estimated differences between the partial derivatives of the "close" function $g(\cdot)$ or the "far" function $h(\cdot)$ before vs. after the occurrence of a tax sale. Estimates are pooled over all years before vs. after a tax sale event. The function $g(\cdot)$ gives a weight to each property within the inner ring equal to 0.1 minus the distance to the tax sale property in miles. $h(\cdot)$ instead adds up with equal weights the number of tax sales within 0.25 miles of the event property. Columns (2), (3), and (4) include a dummy equal to unity for sales of multi-family homes. Columns (3) and (4) include controls for the number of bedrooms, bathrooms, floor space, lot size, and a quadratic in property age. Column (4) winsorizes continuous property controls at the 1st/99th percentile. Each specification includes a full set of tract × year and month fixed effects. Standard errors clustered at the tract-year level in parentheses. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

46.62–54.11 distance-weighted tax sales, the 99.5th to 99.9th percentile is 54.21–63.95, and the 99.9th percentile to the maximum is 64.12–94.11. These positive spillovers remain even after controlling for multi-family property status or a battery of variables capturing property quality (winsorized or unwinsorized).

The magnitude and sign of the estimates indicate an inverted U-shaped relationship between the concentration of tax sales and price externalities. Price spillovers are negative on average, positive if there are many tax sales in a small neighborhood (+2% at the 99th percentile in column 4), but then quickly fall back towards zero if distress is overly prevalent. These large positive effects in tax sale clusters points to the bulk-buying mechanism highlighted in Ganduri, Xiao, & Xiao (2022), who argue that institutional investors help prop up distressed areas by buying adjacent properties slated for redevelopment. Further, there is no evidence of positive spillovers in the top part of the distribution of the "far" function $h(\cdot)$, echoing findings in Pennington (2021) that demand effects of building redevelopment tend to be more hyperlocal than supply effects.

The model in (4.1) can be easily redefined to isolate short-run treatment effects by defining before (B) and after (A) as one year before vs. one year after a tax sale, respectively. To highlight comparability to extant research on mortgage foreclosures, Appendix C shows that tax lien sales reduce market prices of neighboring properties by between 1.5% and 2.3% within a year. The point estimates for the difference $\delta_{C,B} - \delta_{C,A}$ are very close to the -2% very nearby pricing spillover repeatedly replicated elsewhere (e.g. Anenberg & Kung 2014) and originally obtained by Campbell, Giglio, & Pathak (2011) for foreclosures. Interestingly, the positive spillovers observed in areas undergoing waves of tax delinquency do not appear in the short-run after dropping multi-family properties from the sample. Given time to build in converting single-family to multi-family units, this provides further evidence in favor of the commercial redevelopment channel of gentrification through tax sales to institutional investors.

4.2 Spillover Estimates Using Empirical Derivatives

Although the model in (4.1) does account for local time trends with the inclusion of $\alpha_{c,t}$, it relies on strong functional form assumptions and requires the researcher to impose distance thresholds to define a "close" or "far" property. To avoid making parametric assumptions, I also explore how estimates of spillovers of tax sales evolve over time and continuously with distance to the treated property using the semi-parametric empirical derivatives (ED) method of Diamond & McQuade (2019). This method generalizes the ring DD described in Appendix C.1, obviating the need to select inner and outer ring radii while also controlling for very local price trends.[21]

Figure 8 and the accompanying point estimates in Table 4 display results from applying the ED method to D.C. neighborhoods. Panel C indicates that prices of homes sold in a recently gentrifying area jump by 5-7% after an institutional buyer takes possession of a nearby tax-distressed property; the observed spillover effects grow over time and are stronger the closer a house is located to the tax sale property. The impact on neighboring home prices is instead negative in non-gentrifying areas; prices of nearby homes are 5% lower after a tax sale occurs (Panel D). Crucially, Table 4 indicates no statistically significant pre-trends at any distance within a 0.5 mile radius.

Panels A and B offer a placebo test where I categorize neighborhoods based on population flows from 1990-2005 instead of 2005-2019. Compared to Panel C, I observe similar, but weaker pricing trends with respect to time and distance for tracts which were gentrifying prior to the start of my tax sale sample. Moreover, prices evolve similarly between initially gentrifying (Panel A) and non-gentrifying areas (Panel B). In both Panel A and Panel B, the point estimates are wider due to the smaller incidence of gentrifying tracts prior to 2005, and the majority of the point estimates are statistically insignificant. Since the earlier subperiod occurs before the start of my tax lien sample, this points to institutional investors contributing directly to new gentrification rather than merely amplifying the consequences of already settled gentrifiers. Tax sales thus appear to expand the geographic scope of crowd-out of lower-income individuals and increase within-city inequality.

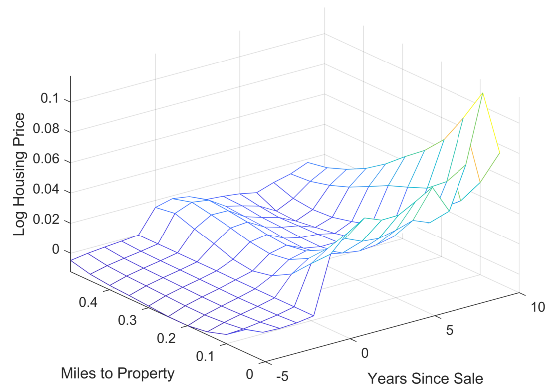
My finding of negative *average* short-run and long-run pricing spillovers for a very different form of property distress hints at similar short-run economic mechanisms, such as peer effects, social stigma, or fire sales, operating on tax sale neighborhoods as in the typical mortgage foreclosure context. To further investigate the underlying economic forces, in Figure 11 I estimate the empirical derivatives pricing surface for tax sale ring neighborhoods split by measures of distress (Panel A) or market thickness (Panel B). In neighborhoods where *ex ante* mortgage foreclosure volume is highest, prices increase over time following a tax sale, with the largest effects exhibited at longer horizons (5-10 years) and close distances. There are no clear post-trends or statistically significant point estimates for low distress neighborhoods containing no mortgage foreclosures in the past year.

The results in Panel A corroborate the idea that the concave pricing pattern with respect to tax sale volumes from the estimates in Table 3 is due to development options having greater value in instances where investors can simultaneously scoop up many neighboring properties. To the extent that foreclosed properties tend to be physically deteriorating (Gerardi et al. 2015; Biswas et al. 2021), this result is also consistent with the blight reduction mechanism emphasized in Ganduri & Maturana (2021), who study the local effects of non-profit property rehabilitation. Because

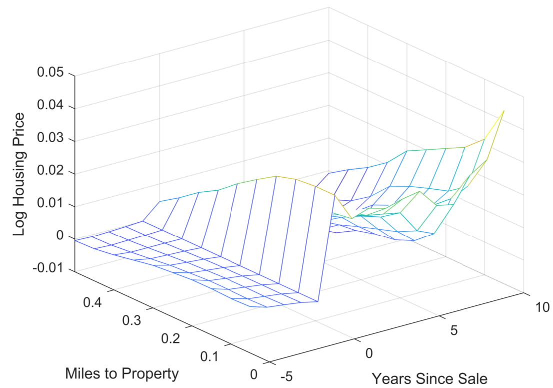
[21] The kernel estimator used to smooth the house price derivatives relies on several tuning parameters that control smoothing along the time and coordinate dimensions, the dimensions of the bowtie drawn around the tax sale event in the middle of the ring, and sample selection. In my baseline estimation, I use the same tuning parameters as in Diamond & McQuade 2019, except I reduce the ring radius from 1.5 miles to 0.5 miles. The general shape of the 3-D pricing surfaces in Figure 8 is insensitive to the choice of all tuning parameters except ring radius – a larger ring radius tends to reduce my point estimates and lead to larger standard errors. See Appendix C for details on the calibration and robustness to tuning parameter choices.

FIGURE 8. Empirical Derivative Estimates of Tax Sale Pricing Spillovers

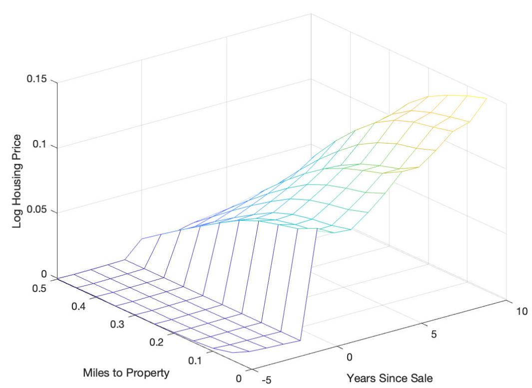
A. Sales in Previously Gentrifying Tracts



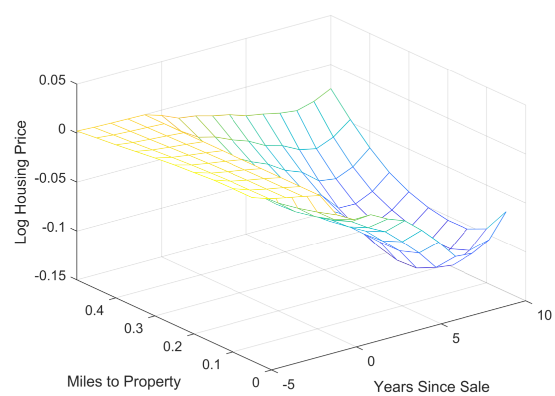
B. Sales in Previously Non-gentrifying Tracts



C. Sales in Recently Gentrifying Tracts



D. Sales in Recently Non-gentrifying Tracts



Notes: Each panel shows the continuous evolution of log sale prices for homes located within a 0.5 mile radius of a property sold at tax lien auction in Washington, D.C. Estimates obtained by applying the empirical derivatives estimator of Diamond & McQuade (2019) and excluding all properties within 0.01 miles of the tax sale property. Tax lien sale properties are excluded from the estimation sample to avoid reverse causality. All estimates are relative to the year before the event of a tax lien sale to an institutional buyer. I identify institutional buyers using the keywords: "LLC", "FUND", "INC", "BANK", "REALTY", "PARTNERS", "CAPITAL", "TRUST", "CORPORATION", "PLLC." Panel A reports estimates for gentrifying tracts, while Panel B does the same for non-gentrifying tracts according to the Census tract type definitions obtained by estimating the population flows model described in Section 3 over the period 1990-2005. Panels C and D conduct the same exercise as in Panels A and B, respectively, except with gentrifying areas identified by running the classification model over the later period in 2005-2019. See Appendix C.3 for details on implementation of the empirical derivatives estimator.

TABLE 4. Pricing Effects of Tax Sales on Nearby Properties in Recently Gentrifying Tracts

Distance (mi.) →	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Event time (years) ↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
−5	0.0012	−0.0013	−0.0018	−0.0017	−0.0017	−0.0012	−0.0013	−0.0012	−0.0013	−0.0010
	(0.0054)	(0.0040)	(0.0036)	(0.0034)	(0.0033)	(0.0031)	(0.0028)	(0.0024)	(0.0019)	(0.0012)
−4	0.0002	−0.0008	−0.0009	−0.0009	−0.0010	−0.0008	−0.0009	−0.0009	−0.0011	−0.0010
	(0.0030)	(0.0025)	(0.0023)	(0.0022)	(0.0021)	(0.0020)	(0.0018)	(0.0015)	(0.0012)	(0.0007)
−3	0.0003	−0.0003	−0.0004	−0.0004	−0.0005	−0.0004	−0.0005	−0.0005	−0.0007	−0.0007
	(0.0019)	(0.0016)	(0.0015)	(0.0014)	(0.0014)	(0.0013)	(0.0012)	(0.0010)	(0.0008)	(0.0005)
−2	0.0003	0.0000	0.0000	−0.0001	−0.0002	−0.0001	−0.0002	−0.0002	−0.0004	−0.0004
	(0.0010)	(0.0009)	(0.0008)	(0.0008)	(0.0008)	(0.0007)	(0.0006)	(0.0005)	(0.0004)	(0.0003)
1	0.0796***	0.0746***	0.0729***	0.0700***	0.0640***	0.0561***	0.0454***	0.0347***	0.0234***	0.0109***
	(0.0145)	(0.0125)	(0.0116)	(0.0110)	(0.0106)	(0.0101)	(0.0091)	(0.0079)	(0.0063)	(0.0042)
2	0.0808***	0.0767***	0.0753***	0.0724***	0.0664***	0.0584***	0.0472***	0.0360***	0.0240***	0.0111***
	(0.0143)	(0.0123)	(0.0115)	(0.0109)	(0.0106)	(0.0101)	(0.0091)	(0.0079)	(0.0063)	(0.0043)
3	0.0899***	0.0850***	0.0827***	0.0791***	0.0727***	0.0641***	0.0522***	0.0401***	0.0272***	0.0135***
	(0.0145)	(0.0125)	(0.0116)	(0.0110)	(0.0106)	(0.0101)	(0.0091)	(0.0079)	(0.0063)	(0.0042)
4	0.0997***	0.0931***	0.0900***	0.0854***	0.0785***	0.0691***	0.0560***	0.0426***	0.0286***	0.0140***
	(0.0143)	(0.0123)	(0.0115)	(0.0109)	(0.0106)	(0.0101)	(0.0091)	(0.0079)	(0.0063)	(0.0043)
5	0.1134***	0.1047***	0.1003***	0.0948***	0.0872***	0.0767***	0.0620***	0.0468***	0.0310***	0.0148***
	(0.0141)	(0.0123)	(0.0114)	(0.0109)	(0.0106)	(0.0101)	(0.0091)	(0.0079	(0.0064)	(0.0043
10	0.1683***	0.1585***	0.1502***	0.1408***	0.1283***	0.1124***	0.0905***	0.0688***	0.0470***	0.0237***
	(0.0187)	(0.0166)	(0.0158)	(0.0153)	(0.0147)	(0.0140)	(0.0123)	(0.0106)	(0.0086)	(0.0059)

Notes: This table presents average log house prices at various distances in event time relative to the one year before a tax lien sale in recently gentrifying tracts, defined as tracts which gentrified over 2005 – 2019 according to the model in Section 3. Note that estimates at event time = 0 correspond to cases where the lien is potentially held within the six-month redemption period where foreclosure by the lienholder is not yet possible, and consequently I suppress these from the table due to a lack of clear interpretation. Estimates computed using the empirical derivative method of Diamond & McQuade (2019) described in the main text and Appendix C.3. Standard errors computed using the block bootstrap method with 500 sample draws, where sampling is carried over neighborhoods corresponding to tax lien properties. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

mortgage and tax foreclosure are disjoint events (cf. Appendix A.3), the notion of distress in Panel A is not simply a sample split based on past tax foreclosure waves.

Panel B of Figure 9 shows that the positive pricing effects of tax sales are concentrated in thin markets, while negative pricing effects are concentrated in thick markets. The sign of these patterns runs counter to theoretical predictions when supply effects of home inventory due distressed sale events dominate. The fact that demand effects predominate in the segment of rings which are in thin markets suggests a limited role for supply effects brought on by forced sales. In thin markets, tax sales bring an opportunity for development and conversion of existing properties that have not changed ownership for a long time.[22] Indeed, based on matched CoreLogic transaction histories, for thin neighborhoods the typical owner's housing tenure in a tax sale property is 14.9 years.

4.3 Estimates under Alternative Counterfactuals

Zooming in at close distances in Figure 8 reveals no statistically significant pre-trend along the time dimension. Yet, homes in neighborhoods with tax delinquencies are likely different on observable and unobservable dimensions from those in non-delinquent areas. To move towards causal estimates of the local spillover effects of tax sales, I adapt the research design from Pollmann (2021), who uses machine learning (ML) methods to select counterfactual areas to compare to actual treated sites. The nationwide coverage and large set of conditioning variables describing housing markets, business activity, and demographics I obtain from merging databases allows me to match areas actually experiencing tax sales to almost identical ones without tax-distressed properties.

The basic idea underlying the ML approach is to extend the notion of propensity score matching (PSM) to a spatial setting. I identify candidate treatment neighborhoods and then estimate the probability that those neighborhoods could contain a tax sale property. As in standard PSM, I then match the closest "false positive" neighborhood to an actual tax sale neighborhood and average the differences across the pairs using the estimator proposed by Chernozhukov et al. (2018) which is robust to small errors in the propensity score estimation. In practice, the procedure involves finely discretizing the latitude/longitude grid, repeatedly drawing candidate locations $\nu_g \sim G(\mathbf{Z_g})$, where the vector $\mathbf{Z_g}$ contains the characteristics of grid cell g, and $\mathbf{Z_g}$ is selected by a neural network (à la Athey et al. 2019). Comparing the resulting point estimates from this spatial PSM exercise to those from the ED method assesses the contributions of idiosyncratic confounders to estimated treatment effects.

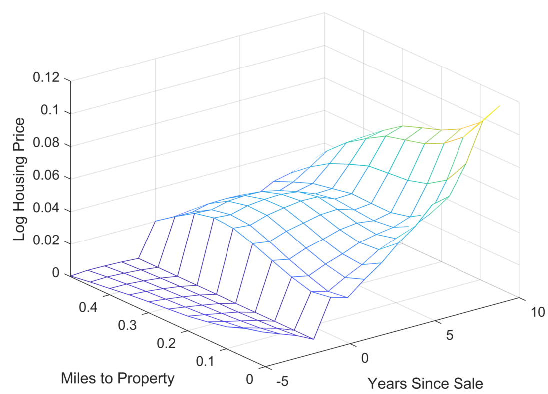
As a final robustness check, I use as a control group properties in neighborhoods around tax-delinquent houses that were originally slated for auction but redeemed before a sale could occur.

[22]Note, thickness is sometimes synonymous with notions of liquidity in housing markets. However, the two are distinct concepts. For instance, consider a neighborhood contained within a top public school district. Such a neighborhood will be in high demand by prospective parents but also have low inventory, leading to low time on market (TOM) when houses are listed for sale. Thus, this market would be thin but relatively liquid. Unfortunately, I cannot construct TOM as a possible measure of liquidity since CoreLogic MLS data are not systematically available prior to the late-2000s when home sale listings were widely published on the internet.

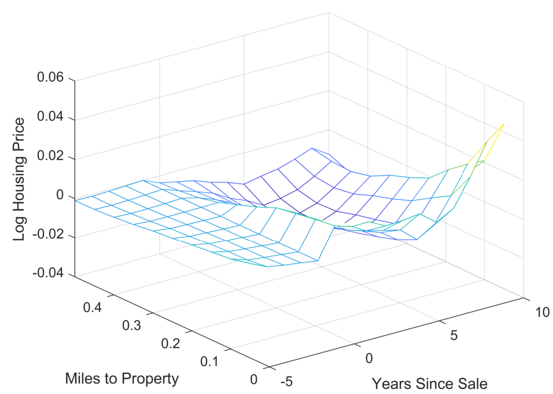
FIGURE 9. Empirical Derivative Estimates by Distress and Market Thickness

A. Tax Sale Neighborhoods by Mortgage Foreclosure Volume

Most Foreclosures

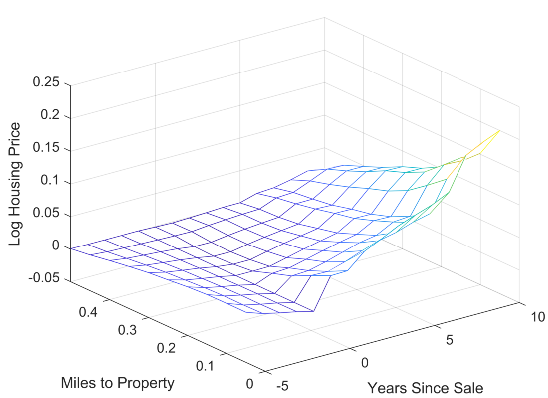


Least Foreclosures

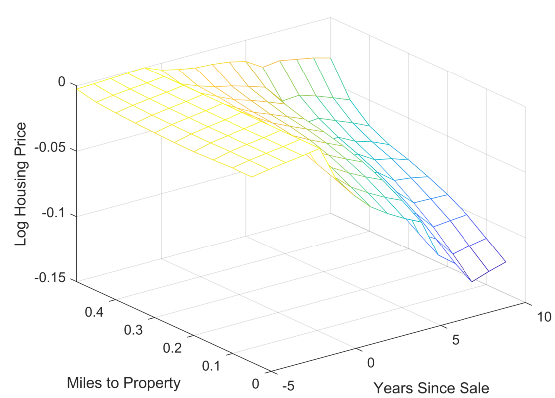


B. Tax Sale Neighborhoods by Thickness

Thinnest Markets



Thickest Markets



Notes: Each panel shows the continuous evolution of log sale prices for homes located within a 0.5 mile radius of a property sold at tax lien auction in Washington, D.C. Estimates obtained by applying the empirical derivatives estimator of Diamond & McQuade (2019) and excluding all properties within 0.01 miles of the tax sale property. Tax lien sale properties are excluded from the estimation sample to avoid reverse causality. All estimates are relative to the year before the event of a tax lien sale to an institutional buyer. I identify institutional buyers using the keywords: “LLC”, “FUND”, “INC”, “BANK”, “REALTY”, “PARTNERS”, “CAPITAL”, “TRUST”, “CORPORATION”, “PLLC.” Panel A reports estimates for most vs. least distressed tax sale neighborhoods, as measured by top quartile vs. below median ring mortgage foreclosure volume in the year before the tax sale occurs. I take below median to mean least distressed because the distribution of foreclosures is very right-skewed, with an annual median of zero within-ring foreclosures but a range of 6-19 foreclosures within the 99th percentile. Panel B instead splits by lowest vs. highest quartile of overall ring transaction volume in the year before a tax sale occurs. See Appendix C.3 for details on implementation of the empirical derivatives estimator.

Identifying this counterfactual group of properties is possible in jurisdictions where the tax authority maintains lists of delinquent bills.[23] Defining the control/placebo group based on the redemption status of the property rather than distance to tax sale event overcomes a key econometric issue underlying the ring strategy: cross-contamination of the estimated effects in cases where two tax sale properties are located very close to each other (Butts 2021a). As shown in the map for D.C. (Figure 1), tax lien purchases by institutional buyers tend to be highly geographically clustered. I present the full results using the ML methods and just-redeemed set of properties to form control neighborhoods in Appendix C.

5 Do Tax Sales Drive Local Demographic Change?

The results in the preceding section imply institutional real estate investment through local tax auctions contributed to increasing income-based sorting among homeowners within cities in recent years. Have these investment behaviors also contributed to demographic change in neighborhoods experiencing tax delinquency cases? If delinquent taxpayers are driven from their homes through tax foreclosure, what can we say about the race of the homeowners who move into these now more expensive neighborhoods in their place? Answering these questions is of great importance for evaluating the benefits and costs of proposed reforms to local tax collection systems, given extant evidence of a racial wealth gap arising from disparities in homeownership (Perry et al. 2018) and housing returns (Kermani & Wong 2021).

5.1 Addressing Empirical Challenges

I face two major empirical obstacles in examining how tax sales might alter the demographic composition of homeowners. The first is that the race of buyers and sellers is not readily observable in most real estate transaction databases. Race is also not reported for the owners and investors in the separate set of tax sale data I collect in this paper. Researchers have tried to impute racial identities by either (i) merging the property records of interest to Home Mortgage Disclosure Act (HMDA) data from loan applications, or (ii) estimating the likelihood a counterparty belongs to a racial/ethnic category by matching surnames to self-reported race in the Census Surname Lists. I do not pursue the former option because the vast majority of tax sales do not have a mortgage attached (cf. Appendix A.3), and given recent evidence of discrimination in mortgage approvals (Giacoletti, Heimer, & Yu 2021), racial demographics are likely to differ substantially between CoreLogic/Zillow ZTRAX and HMDA.

Following other papers in the housing literature (e.g. Humphries et al. 2019; Christensen & Timmins 2021), I use the `wru` package developed by Imai & Khanna (2016) to classify buyers. The

[23] In Appendix E, I provide an example of the tax bills I use to construct a panel of taxpayer-property ids and identify delinquency events resulting in a redemption. Many of these redemption events occur because either the mortgage lender or a family member steps in to protect their equity stake in the house.

package uses Bayes's rule to estimate the homeowner's probability of belonging to a racial category conditional on a surname matched to the Census Surname List and the property address location.[24] Figure 3 summarizes the Bayesian imputation of homeowner race in the CoreLogic Tax data and my tax lien sale data. I show results using both the `wru` probability estimate itself and dummies given the evidence in Figure 3 that dummies are highly sensitive to the assumption of an imposed threshold cutoff.

The second empirical challenge deals with how to identify the causal impact of tax sales on the identity of new homeowners. As noted in the preceding section, removing the confounding influence of very local time trends is key to uncovering these effects. Yet, unlike most previous studies in the mortgage foreclosure spillover literature and my analysis in Figure 5, I am now interested in a discrete variable – the probability that a buyer or seller belongs to a self-identified racial category – rather than a continuous one like sale price. Many neighborhoods are highly segregated and therefore feature "corner solutions" wherein the fraction of white homeowners is either very close to zero or very close to unity. This renders it difficult to apply the empirical derivatives method to non-parametrically study how racial sorting evolves with respect to both time and distance to a tax sale event.[25]

Inspired by Pennington (2021) who quantifies displacement around new market-rate and affordable housing construction, I adopt a semi-parametric version of the empirical derivatives estimator. I estimate regressions of the following form over a set of property transactions with finely binned tax sale radius dummies:

$$URM_{i,r,t} = \sum_{k=-3}^{+5} \left\{ \sum_{d=0.05 \text{ mi}}^{0.5 \text{ mi}} \beta_{d,k}^{close,a} \cdot TaxSale\_Close_{i,t,d,k}^{a} \right. \tag{5.1}$$

$$\left. + \sum_{d=0.5 \text{ mi}}^{1 \text{ mi}} \beta_{d,k}^{far,a} \cdot TaxSale\_Far_{i,t,d,k}^{a} \right\} + \alpha_{r,t} + \gamma_m + \eta' \cdot \mathbf{X_{i,t}} + \nu_{i,r,t}$$

where the outcome is the `wru`-implied probability that property transaction i located in tax sale ring r occurring at time t involves a Black or non-white Hispanic counterparty (either the buyer or seller). I include a set of fine distance bins d where the increments are 0.05 miles within the "close" ring of 0.5 mile radius around a tax sale concluding in event year $k = 0$, and the increments are 0.1 miles outside a 0.5 mile ring around a tax sale. These choices for the "close" and "far" cutoff radii conform to the 0.5 mile search radius I adopt in the empirical derivatives estimator (see Table C.2

[24] More formally, the `wru` package infers a racial category $\mathcal{R}$ using location ℓ and surname $\mathcal{S}$ in the Census surname list, then estimates $\widetilde{p} \equiv \Pr(R_i = \mathcal{R} | L_i = \ell, S_i = \mathcal{S})$ via Bayes's rule. For localities such as D.C. where voter records with racial identities are available, the package can also condition on such information to improve the precision of the estimates. I check that the results go through for three definitions of the race probability: (i) set $URM = 1$ if $\widetilde{p} > 0.5$ for Black or Hispanic, (ii) set $URM = 1$ if highest probability race is Black or Hispanic (i.e. exactly what Imai & Khanna (2016) do), and (iii) $\widetilde{p}$ itself.

[25] I discuss in Appendix C.3 why the solution algorithm for the empirical derivatives estimator cannot easily accommodate binary outcomes.

for the calibration of that method).[26] The $\alpha_{r,t}$ represent a full set of ring $\times$ year fixed effects, and γ_m are transaction month dummies. The vector $\mathbf{X_{i,t}}$ includes property-level controls such as latitude and longitude, the number of bathrooms and bedrooms, floor space, lot size, and a quadratic in building age. Adding property controls allows for race-specific demand for local amenities, as in the decomposition of population growth conducted by Baum-Snow & Hartley (2020).[27]

The variables $TaxSale\_Close^a_{i,t,d,k}$ and $TaxSale\_Far^a_{i,t,d,k}$ measure exposure to tax sales. To fix ideas, consider the case where $TaxSale\_Close^a_{i,2005,0.1,-1} = 1$. This refers to a scenario where property i is purchased in 2005 within 0.1 miles of a tax sale property eventually transferred in 2006 to an investor of type a. I include separate interactions of $TaxSale\_Close$ and $TaxSale\_Far$ with dummies for institutional or individual lien investors, with the types indexed by superscript a, to account for differential effects on neighborhood composition depending on the investment strategy. To maintain a parsimonious number of dummies, I bin the endpoints at $k = -3$ and $k = +5$, so that the $TaxSale^a_{i,t,d,-3}$ and $TaxSale^a_{i,t,d,+5}$ dummies are equal to one if the transaction occurs at any point 3 or more years ahead of, or 5 or more years after a tax sale within d miles, respectively. I include a separate set of dummies for faraway tax sales to account for any spatial spillovers from non-contiguous neighborhoods which also experience tax sales.

The $\beta^{close,a}_{d,k}$ coefficients in equation (5.1) are difficult to directly interpret to the extent that some properties are contained within overlapping tax sale rings. I also estimate the following pooled OLS (POLS) specification which collapses the event study model to the property level:

$$URM_{i,c,t} = \sum_{d=0.05 \text{ mi}}^{0.5 \text{ mi}} \beta^{close,a}_{d} \cdot TaxSale\_Close^a_{i,t,d} \tag{5.2}$$

$$+ \sum_{d=0.5 \text{ mi}}^{1 \text{ mi}} \beta^{far,a}_{d} \cdot TaxSale\_Far^a_{i,t,d} + \alpha_{c,t} + \gamma_m + \eta' \cdot \mathbf{X_{i,t}} + e_{i,r,t}$$

The POLS model collapses over the k leads and lags in (5.1) by taking a 5-year sum of tax sales occurring in event years $1 \leq k \leq 5$. Therefore, $TaxSale\_Close^a$ and $TaxSale\_Far^a$ are now the *cumulative* exposure of property transaction i to tax sales to investor type a occurring within the last five years at distances within 0.5 miles or outside a 0.5 mile radius, respectively. For instance, $TaxSale\_Close^a_{i,2010,0.1} = 3$ would refer to the case where property i is sold in 2010 and lies within 0.1 miles of three tax sale transfers completed some time between 2006 and 2009.

[26] I discretize event time by first computing the number of days ℓ between the transaction date and the lien auction date + the end of the redemption period. For transactions occurring before the lien sale event ($\ell < 0$) I then round down to the nearest year, while for transactions occurring after the lien sale event ($\ell > 0$) I round up to the nearest year. I discretize event time in this fashion because the bulk of tax lien sales occur at annual sales (cf. Table 1), and so the events effectively occur at an annual frequency. It also limits computational intensity by retaining a smaller set of dummies to be estimated in (5.1).

[27] Controlling for the geo-coordinates of a property operates like the control function in a spatial regression discontinuity design which allows the outcome to change continuously over location in space, much like the empirical derivatives strategy.

The pooled specification has the advantage of being more statistically powered given the much lower number of dummies to be estimated (collapsing from $8 \times 15 = 120$ $\widehat{\beta}_{d,k}$ to 15 $\widehat{\beta}_d$ coefficients). It is also easier to directly interpret the coefficients, as the dummies $\beta_d^{close,a}$ quantify the effect on buyer or seller probabilities of an additional nearby tax sale. The POLS specification is akin to the foreclosure wave regression in (4.1), but more general in that it allows for non-linearities with respect to distance to tax sale properties. I report the POLS results in Appendix C.5. The results are qualitatively similar across the event study and POLS specifications, although the negative effects of tax sales to institutional investors on underrepresented buyer and seller probabilities are more localized when accounting for overlapping rings and pooling across time periods.

5.2 Tax Sales & The Racial Composition of Housing Transactions

Figure 10 plots the 3-D surface representation of the event study coefficients $\beta_{d,k}^{close,a}$ estimated from equation (5.1) by applying two definitions of the buyer race probability dummy URM_buyer and running separate models for rings associated with institutional vs. individual buyers of tax claims. Figure 11 shows the event study coefficients with 95% confidence interval bars corresponding to the Panel B estimates in Figure 10. While the 3-D surface trends are consistent across the institutional and individual tax sale rings – underrepresented minority buyer probabilities trend slightly upward before dropping after a tax sale event – there are clear differences in magnitude and statistical significance. For the continuous probability definition (Panel B), the probability surface is statistically flat for individual investors. In contrast, in neighborhoods defined by tax sales to institutional investors, underrepresented buyer probabilities are largely flat until dropping by 1-2 p.p. after three years following a tax sale (Panel A), or dropping by 3-4 p.p. using the continuous probability measure (Panel B).

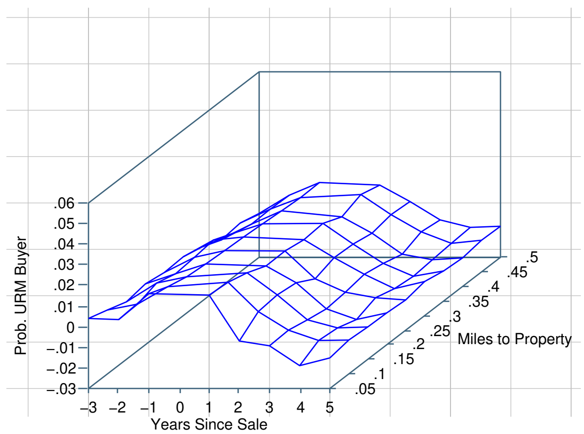
These are economically large effects compared to the average underrepresented buyer probability across all observations at $k = -1$ of 12.4% using the dummy definition, or 23.2% using the continuous definition. The estimated effects for purchases in the post-tax sale period are slightly stronger at close distances, but are statistically flat within a 0.5 mile ring radius. Moreover, consistent with the preceding evidence that institutional investors are somewhat more likely to buy tax liens in already gentrifying areas, the underrepresented buyer probabilities are slightly lower than for individual investors in the pre-period.

Figure 3 documents that tax sale properties are more likely to be owned by Black or Hispanic taxpayers than the general population of single family homes. There are other, more indirect reasons why displacement of racial minorities in tax sale neighborhoods might occur besides through tax foreclosures themselves. For instance, if the local cost of living increases as tax assessed values increase, then liquidity constrained homeowners may be induced to sell and move to lower-cost areas, especially if the expected capital gain is higher due to the spillover effects of redevelopment (He 2022). Homophily, or a desire to live among members of one's own racial group, could encourage non-white homeowners to voluntarily leave as white buyers enter (Ihlanfeldt & Scafidi 2002).

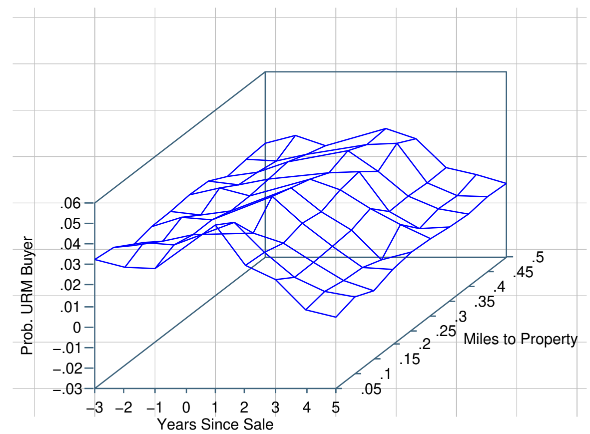
FIGURE 10. 3-D Event Study Estimates of Underrepresented Minority Buyer Probabilities

A. Underrepresented Minority Buyer Dummy

Institutional Investors

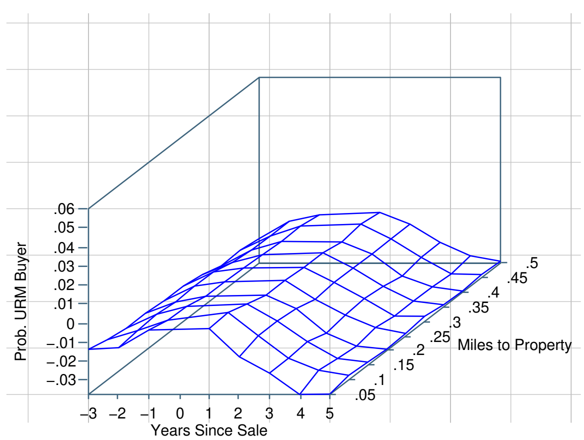


Individual Investors

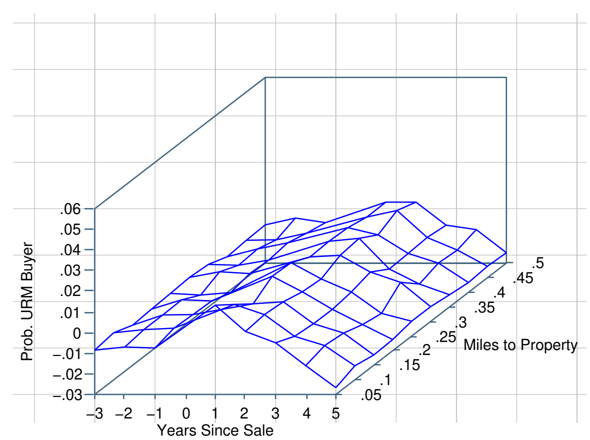


B. Continuous Underrepresented Minority Buyer Probability

Institutional Investors



Individual Investors

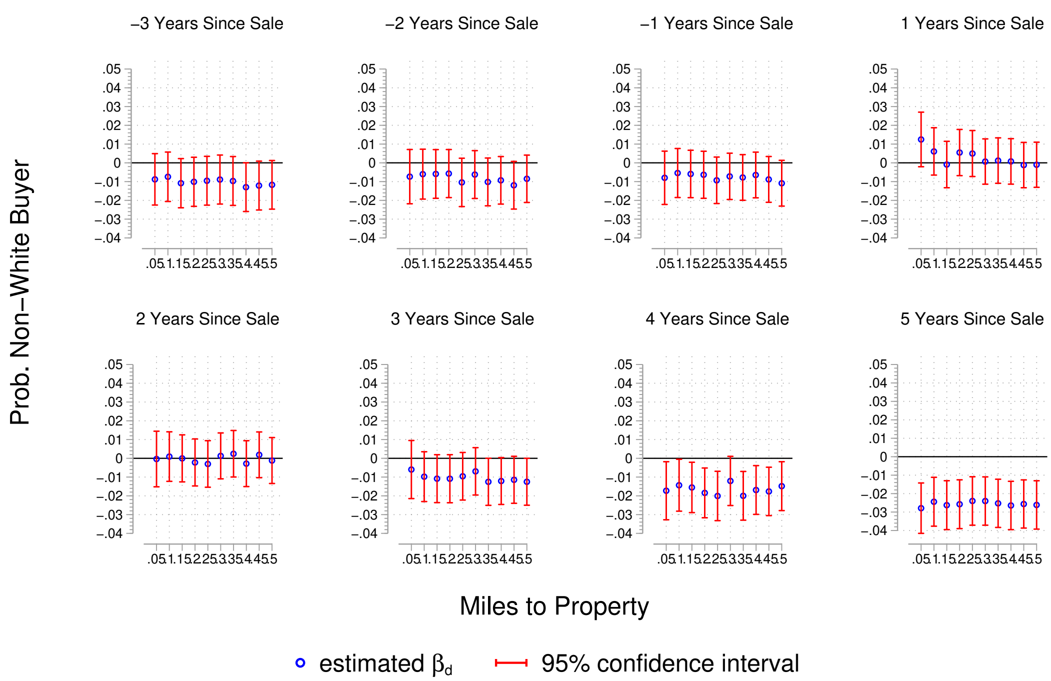


Notes: Each panel shows the evolution of the probability that a homebuyer is non-white and non-Asian for transactions located within a 0.5 mile radius of a property sold at tax lien auction in Washington, D.C. Estimates obtained by applying the semi-parametric event study model in equation (5.1) and excluding all properties within 0.01 miles of the tax sale property. Tax lien sale properties are excluded from the estimation sample to avoid reverse causality. All estimates are relative either to the event of a tax lien sale to an institutional buyer (left panels) or an individual buyer (right panels). I identify institutional buyers using the keywords: "LLC", "FUND", "INC", "BANK", "REALTY", "PARTNERS", "CAPITAL", "TRUST", "CORPORATION", "PLLC." Individuals exclude institutional buyers and non-profit buyers identified using the keywords: "PRAYER", "CHURCH", "COMMUNITY", "FAITH", "UNIVERSITY", "COLLEGE", "SCHOOL", "BAPTIST", "FOUNDATION", "GOVERNMENT", "EMBASSY", "CENTER", "COOPERATIVE", "FRIENDSHIP", "MINISTRIES", "FEDERAL", "REHABILITATION." Panel A reports estimates using a 0/1 dummy constructed by setting the variable equal to 1 if the race of the buyer is likely to be Black or Hispanic, as measured by an imputed probability >50% for either of those two races; set equal to 0 if the probability >50% for any of the other races, and set to missing if no race has a 50% majority. Panel B conducts the same exercise except using a continuous probability estimated via Bayes's rule and following Imai & Khanna (2016).

FIGURE 11. 2-D Event Study Estimates of Underrepresented Minority Buyer Probabilities

A. Institutional Investors

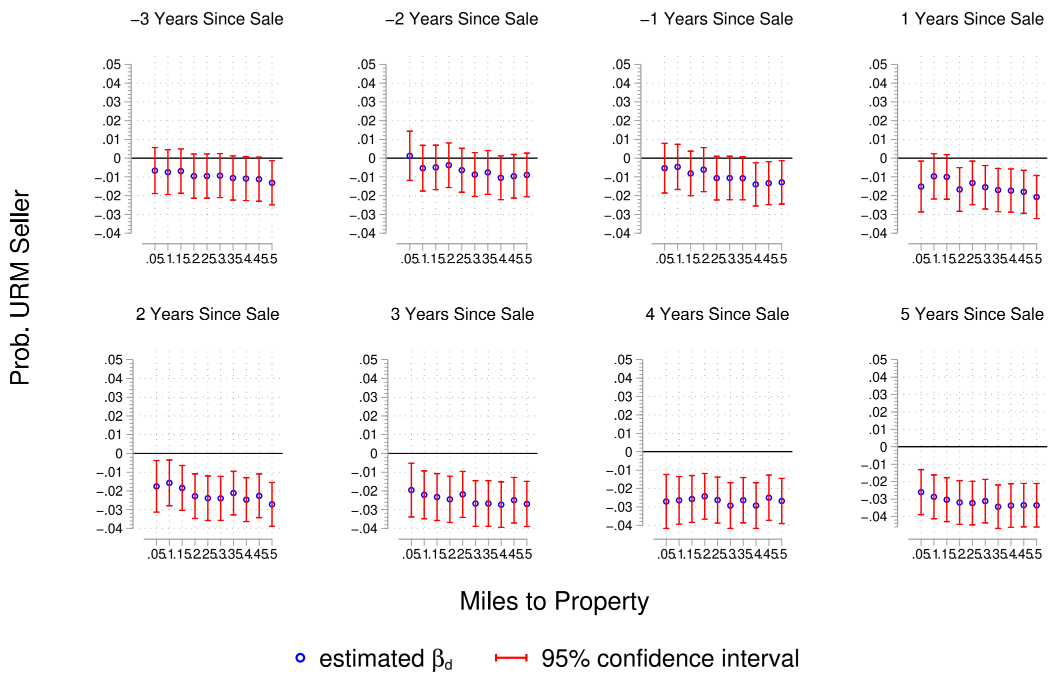
B. Individual Investors



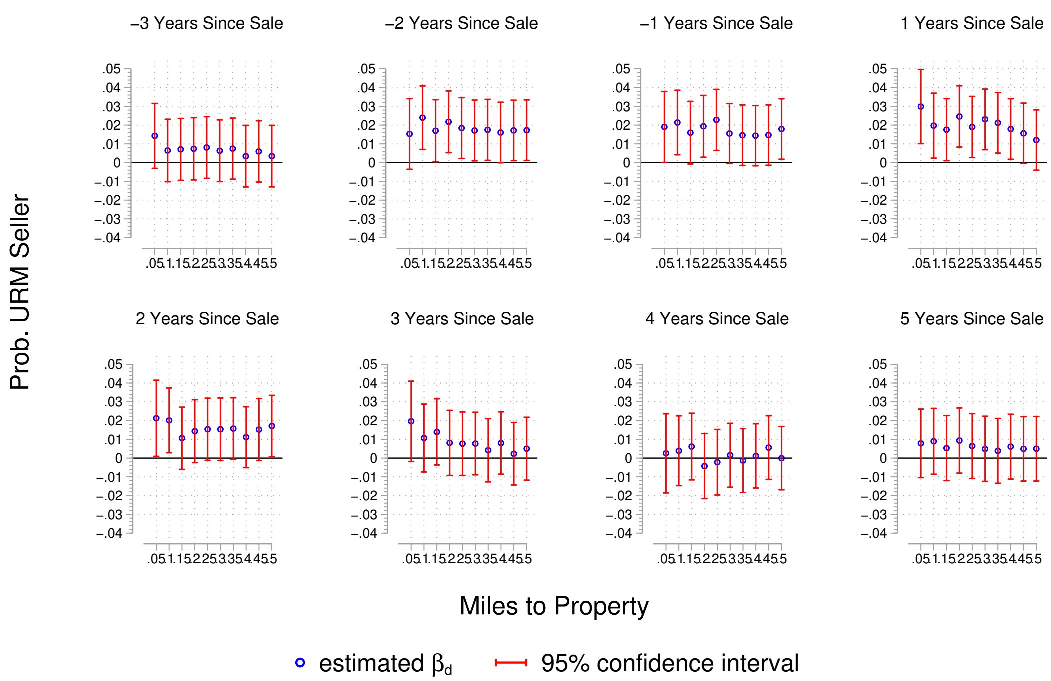
Notes: Each panel shows the point estimates with 95% confidence interval bars corresponding to the 3-D event study surfaces plotted in Panel B of Figure 10 for the continuous non-white and non-Asian homebuyer probability measure. I cluster standard errors at the tax sale ring level. Estimates are relative to the year of the tax sale to avoid overlap between the lien redemption/foreclosure period and surrounding home transaction dates. See text for details.

FIGURE 12. 2-D Event Study Estimates of Underrepresented Minority Seller Dummies

A. Institutional Investors



B. Individual Investors



Notes: Each panel shows the point estimates with 95% confidence interval bars corresponding to the estimates from equation (5.1) for either institutional (Panel A) or individual (Panel B) tax lien investors. I cluster standard errors at the tax sale ring level. Estimates are relative to the year of the tax sale to avoid overlap between the lien redemption/foreclosure period and surrounding home transaction dates. See text for details.

Further, as white buyers move in, the composition of surrounding businesses changes (Glaeser, Luca, & Moszkowski 2020). Self-segregation could then also be driven by race-specific preferences over neighborhood amenities, as highlighted by Waldfogel (2008) for restaurants.

If there is displacement of racial minority homeowners, then in the data one should observe that that the share of non-white home sellers *increases* around a tax sale event. To test this implication, Figure 12 repeats the event study analysis of equation (5.1) using the underrepresented minority status of the *seller* rather than the buyer as the outcome. Contrary to the displacement hypothesis, the patterns are similar to those in Figure 11. For tax sales to institutional investors, underrepresented minority seller probabilities are flat at close distances prior to a tax sale, but decline by 3 p.p. within five years. This is a large decline relative to the average underrepresented seller probability in the year before a tax sale of 10.2%. For liens sold to individual investors, the time gradient in the fraction of minority sellers is statistically flat due to wide confidence intervals, but with a slight positive pre-trend. As individual investors are more likely to purchase tax claims in non-gentrifying tracts, this could reflect homeowners in low-income concentration tracts, who are more likely to be non-white, looking to move as homes in the area become more distressed.

Overall, my results support arguments in the popular media that tax sales to institutional investors lead to racial demographic change through property redevelopment, with subsequent buyers and sellers in a neighborhood less likely to be Black or Hispanic, even after controlling for very local time trends. However, some incumbent minority homeowners choose to stay in formerly tax-distressed neighborhoods and benefit from the appreciation in property values. While tax sales in the U.S. amplify gentrification by transferring wealth from delinquent homeowners to local governments and investors, they also entail significant transfers between neighbors who share demographic characteristics. Thus, proposed reforms to the tax sale system – such as earmarking funds for redemption assistance programs or caps on penalties (Rao 2012) – may be subject to not only standard equity-efficiency tradeoffs, but also to tradeoffs between fostering vertical equity at the cost of increased horizontal inequality in housing wealth.

6 Conclusion

I study the spillovers effects on local prices and demographics of institutional investors entering the residential real estate market through tax-distressed property auctions. I compile a new dataset from public records of tax lien and tax deed sales in major metro areas to present several stylized facts about this understudied segment of the housing market. Tax foreclosure sales are markedly underpriced relative to mortgage foreclosures, typically selling for under 10% of the home's previously assessed market value. Compared to individual or non-profit investors, institutional investors are more likely to act as opportunistic developers who invest in claims to homes located in currently gentrifying areas or neighborhoods bordering gentrifying areas. Moreover, the tax lien auction circuit is dominated by a small handful of amorphous LLCs, which in many cases act as intermediaries for large private real estate funds, and which earn high returns regardless of whether they ultimately foreclose on a property due to high statutorily guaranteed interest rates.

Using unusually detailed data from Washington, D.C. as a case study, I document small, immediate negative price effects, on average, of tax lien sales on neighboring properties. However, in both previously gentrifying and currently gentrifying areas, large positive pricing spillovers emerge within three years for tax liens purchased by institutional investors. Accompanying these positive local pricing effects is a decline in subsequent housing purchases by underrepresented minority buyers, driven by conversion of former tax sale properties into luxury multi-family buildings. In recent years, private capital's implicit partnership with local tax authorities to recover lost tax revenues has amplified gentrification and the Black-white wealth gap through displacement of incumbent homeowners. My results point to alleviating municipal financing constraints – potentially by conditioning federal and state revenue reallocation on the results of municipal balance sheet stress tests – as a way to combat within-city wealth inequality and mitigate the crowd-out of low-income, elderly residents from high cost of living areas.

References

Albouy, D. (2016): "What Are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities," *Review of Economics and Statistics*, 98(3): 477–487.

Alexander, F. S. (2000): "Tax Liens, Tax Sales, and Due Process," *Indiana Law Journal*, 75(3): 747-808.

Alm, J., T. Hodge, G. Sands, & M. Skidmore (2014): "Property Tax Delinquency - Social Contract in Crisis: The Case of Detroit," Victoria Business School Working Paper, No. 3149.

Alm, J., Z. Hawley, J. M. Lee, & J. Miller (2016): "Property Tax Delinquency and its Spillover Effects on Nearby Properties," *Regional Science and Urban Economics*, 58: 71-77.

Amornsiripanitch, N. (2021): "Why Are Property Tax Rates Regressive?" *mimeo*, Yale.

Anenberg, E. & E. Kung (2014): "Estimates of the Size and Source of Price Declines Due to Nearby Foreclosures," *American Economic Review*, 104(8): 2527-2551.

Asquith, B.J., E. Mast, & D. Reed (2021): "Local Effects of Large New Apartment Buildings in Low-Income Areas," forthcoming, *Review of Economics and Statistics.*

Athey, S., G. W. Imbens, J. Metzger, and E. M. Munro (2019): "Using Wasserstein Generative Adversarial Networks for the Design of Monte Carlo Simulations," NBER Working Paper Series, No. 26566.

Atuahene, B. & C. Berry (2019): "Taxed Out: Illegal Property Tax Assessments and the Epidemic of Tax Foreclosures in Detroit," *UC Irvine Law Review*, 9(4): 847-886.

Avenancio-León, C. & T. Howard (2022): "The Assessment Gap: Racial Inequalities in Property Taxation," *Quarterly Journal of Economics*, 137(3): 1383-1434.

Bartlett, R., A. Morse, R. Stanton, & N. Wallace (2022): "Consumer-Lending Discrimination in the FinTech Era," *Journal of Financial Economics*, 143(1): 30-56.

Baskett, C.M. & C.G. Bradley (2021): "Property Tax Privateers," *Virginia Tax Review*, 41(1): 89-135.

Baum-Snow, N. & D. Hartley (2020): "Accounting for Central Neighborhood Change, 1980-2010," *Journal of Urban Economics*, 117: 103228.

Bayer, P., H. Fang, & R. McMillan (2014): "Separate When Equal? Racial Inequality and Residential Segregation," *Journal of Urban Economics*, 82: 32-48.

Bayer, P., C. Geissler, K. Mangum, & J.W. Roberts (2020): "Speculators and Middlemen: Strategy and Performance of Investors in the Housing Market," *Review of Financial Studies*, 33(11): 5212-5247.

Berry, C. (2018): "Estimating Property Tax Shifting Due to Regressive Assessments," University of Chicago Harris School of Public Policy and the Center for Municipal Finance Research Brief, 1-11.

Biswas, A., C. Cunningham, K. Gerardi, & D. Sexton (2021): "Foreclosure Externalities and Vacant Property Registration Ordinances," *Journal of Urban Economics*, 123.

Bhutta, N. & A. Hizmo (2021): "Do Minorities Pay More for Mortgages?" *Review of Financial Studies*, 34(2): 763-789.

Brueckner, J.K. & S.S. Rosenthal (2009): "Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns Be Rich?" *Review of Economics and Statistics*, 91(4): 725-743.

Butts, K. (2021a): "Difference-in-Differences Estimation with Spatial Spillovers," *mimeo*, University of Colorado Boulder.

Butts, K. (2022): "Difference-in-Differences with Geocoded Microdata," forthcoming, *Journal of Urban Economics: Insight*.

Campbell, J., S. Giglio, & P. Pathak (2011): "Forced Sales and House Prices," *American Economic Review*, 101(5): 2108-2131.

Case, K.E. & R.J. Shiller (1987): "Prices of Single Family Homes Since 1970: New Indices for Four Cities," NBER Working Paper, No. 2393.

Case, K.E. & R.J. Shiller (1989): "The Efficiency of the Market for Single-Family Homes," *American Economic Review*, 79(1): 125-137.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, & J. Robins (2018), "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21(1): C1–C68.

Chinco, A. & C. Mayer (2016): "Misinformed Speculators and Mispricing in the Housing Market," *Review of Financial Studies*, 29(2): 486-522.

Christensen, P. & C. Timmins (2021): "Sorting or Steering: Experimental Evidence on the Economic Effects of Housing Discrimination," forthcoming, *Journal of Political Economy*.

Couture, V. & J. Handbury (2019): "Urban Revival in America, 2000 to 2010," NBER Working Paper, No. 24084.

Deerson, David, Joshua Polk, & Christina Martin (2023): "Home Equity Theft: Tax Foreclosure Laws in 50 States and the District of Columbia," Pacific Legal Foundation report: https://homeequitytheft.org/appendix-laws.

Derenoncourt, E., C.H. Kim, M. Kuhn, & M. Schularick (2021): "The Racial Wealth Gap, 1860-2020," *mimeo*, Princeton.

Diamond, R. & T. McQuade (2019): "Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Development," *Journal Political Economy*, 127(3): 1063-1117.

Ding, L. & J. Hwang (2020): "Effects of Gentrification on Homeowners: Evidence from a Natural Experiment," *Regional Science and Urban Economics*, 83: 1-14.

Ding, L. & J. Hwang, & E. Diviringi (2016): "Gentrification and Residential Mobility in Philadelphia," *Regional Science and Urban Economics*, 61: 38-51.

Fisher, L.M., L. Lambie-Hanson, & P. Willen (2015): "The Role of Proximity in Foreclosure Externalities: Evidence from Condominiums," *American Economic Journal: Economic Policy*, 7(1): 119-140.

Ganduri, R. & G. Maturana (2021): "Do Property Rehabs Affect Neighboring Property Prices?" *mimeo*, Emory.

Ganduri, R., S.C. Xiao, & S.W. Xiao (2022): "Tracing the Source of Liquidity for Distressed Housing Markets," forthcoming, *Real Estate Economics.*

Ganong, P. & D. Shoag (2017): "Why Has Regional Income Convergence in the U.S. Declined?" *Journal of Urban Economics*, 102: 76-90.

García, D. (2022): "Second-home Buying and the Housing Boom and Bust," *Real Estate Economics*, 50(1): 33-58.

Gargano, A. & M. Giacoletti (2022): "Cooling Auction Fever: Evidence from the Housing Market," *mimeo*, University of Houston.

Garriga, C., P. Gete, & A. Tsouderou (2020): "Investors and Housing Affordability," St. Louis Fed Working Paper, No. 2020-047A.

Gaubert, C., P.M. Kline, D. Vergara, & D. Yagan (2021): "Trends in U.S. Spatial Inequality: Concentrating Affluence and a Democratization of Poverty," NBER Working Paper, No. 28385.

Gerardi, K., E. Rosenblatt, P.S. Willen, & V. Yao (2015): "Foreclosure Externalities: New Evidence," *Journal of Urban Economics*, 87: 42-56.

Gerardi, K., P. Willen, D.H. Zhang (2021): "Mortgage Prepayment, Race, and Monetary Policy," forthcoming, *Journal of Financial Economics.*

Giacoletti, M., R.Z. Heimer, & E.G. Yu (2021): "Using High-Frequency Evaluations to Estimate Discrimination: Evidence from Mortgage Loan Officers," *mimeo*, USC.

Glaeser, E.L., H. Kim, & M. Luca (2018): "Nowcasting Gentrification: Using Yelp Data to Quantify Neighborhood Change," *AEA Papers and Proceedings*, 108: 77-82.

Glaeser, E.L., M. Luca, & E. Moszkowski (2020): "Gentrification and Neighborhood Change: Evidence from Yelp," NBER Working Paper, No. 28271.

Guerrieri, V., D. Hartley, & E. Hurst (2013): "Endogenous Gentrification and Housing Price Dyanmics," *Journal of Public Economics*, 100: 45-60.

Gupta, A. (2019): "Foreclosure Contagion and the Neighborhood Spillover Effects of Mortgage Defaults," *The Journal of Finance*, 74(5): 2249-2301.

Gupta, A., C. Hansman, & P. Mabille (2022): "Financial Constraints and the Racial Housing Gap," *mimeo*, NYU Stern.

Harding, J.P., E. Rosenblatt, & V.W. Yao (2012): "The Foreclosure Discount: Myth or Reality?" *Journal of Urban Economics*, 71: 204-218.

Hayashi, A.T. (2020): "Countercyclical Property Taxes," *University of Virginia, Public Law and Legal Theory Research Paper Series*, No. 2020-28.

He, M. (2022): "Property Tax and Residential Choices," PhD dissertation, University of Chicago.

Hochstenbach, C. & W.P.C. van Gent (2015): "An Anatomy of Gentrification Processing: Variegating Causes of Neighborhood Change," *Environment and Planning*, 47: 1480-1501.

Hodge, T.R., D. P. McMillen, G. Sands, & M. Skidmore (2017): "Assessment Inequity in a Declining Housing Market: The Case of Detroit," *Real Estate Economics*, 45(2): 237-258.

Hodge, T. R (2019): "Decreasing Delinquency through Assessment Reductions: Evidence from Detroit," *Journal of Urban Affairs*, Special Issue: 1-17.

Humphries, J.E., N.S. Mader, D.I. Tannenbaum, & W.L. van Dijk (2019): "Does Eviction Cause Poverty? Quasi-Experimental Evidence from Cook County, IL," NBER Working Paper, No. 26139.

Hwang, J. & L. Ding (2020): "Unequal Displacement: Gentrification, Racial Stratification, and Residential Destinations in Philadelphia," *American Journal of Sociology*, 126(2): 354-406.

Ihlanfeldt, K.R. & B. Scafidi (2002): "Black Self-Segregation as a Cause of Housing Segregation: Evidence from the Multi-City Study of Urban Inequality," *Journal of Urban Economics*, 51: 366-390.

Imai, K. & K. Khanna (2016): "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records," *Political Analysis*, 24: 263-272.

Institute on Metropolitan Opportunity (2019): "American Neighborhood Change in the 21st Century", University of Minnesota Law School, https://www.law.umn.edu/sites/law.umn.edu/files/metro-files/american_neighborhood_change_in_the_21st_century_-_full_report_-_4-1-2019.pdf

Kahn, M.E. (2021): "Racial and Ethnic Differences in the Financial Returns to Home Purchases from 2007 to 2020," NBER Working Paper, No. 28759.

Kahrl, A. (2017): "Investing in Distress: Tax Delinquency and Predatory Tax Buying in Urban America," *Critical Sociology*, 43(2): 199-219.

Kermani, A. & F. Wong (2021): "Racial Disparities in Housing Returns," NBER Working Paper, No. 29306.

Kerstein, R. (1990): "Stage Models of Gentrification: An Examination," *Urban Affairs Quarterly*, 25(4): 620-639.

Khanna, K., K. Imai, & H. Jin (2017): "Who Are You? Bayesian Prediction of Racial Category Using Surname and Geolocation," Technical report, The Comprehensive R Archive Network.

Lambie-Hanson, L., W. Li, & M. Slonkosky (2022): "Real Estate Investors and the U.S. Housing Recovery," *Real Estate Economics*, 50(6): 1425-1461.

Linden, L. & J.E. Rockoff (2008): "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 98(3): 1103-1127.

Loftis, L.B. (2007): *Profit by Investing in Real Estate Tax Liens*, Second Edition, Kaplan: New York, NY.

Marcuse, P. (1985): "Gentrification, Abandonment, and Displacement: Connections, Causes, and Policy Responses in New York City," *Journal of Urban and Contemporary Law*, 28(1): 195-240.

Marcuse, P. (1986): "Abandonment, Gentrification and Displacement: The Linkages in New York City," in *Gentrification of the City*, Eds. Smith N. & P. Williams, Unwin Hyman: London, 153–177.

Pennington, K. (2021): "Does Building New Housing Cause Displacement?: The Supply and Demand Effects of Construction in San Francisco," *mimeo*, UC Berkeley.

Perry, A., J. Rothwell, & D. Harshbarger (2018): "The Devaluation of Assets in Black Neighborhoods," Brookings Metropolitan Policy Program, https://www.brookings.edu/research/devaluation-of-assets-in-black-neighborhoods/.

Pollmann, M. (2021): "Causal Inference for Spatial Treatments," *mimeo*, Stanford.

Rao, J. (2012): "The Other Foreclosure Crisis: Property Tax Lien Sales," NCLC National Consumer Law Center, https://www.nclc.org/issues/the-other-foreclosure-crisis.html.

Schelling, T.C. (1971): "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1: 143–186.

Schelling, T.C. (1978): *Micromotives and Macrobehavior*, Norton: New York, NY.

Schmidt, M. (2021): "Cook County Scavenger Sale Evaluation," University of Chicago Harris School of Public Policy and the Center for Municipal Finance Report, 1-64.

Small, M.L. & M. McDermott (2006): "The Presence of Organizational Resources in Poor Urban Neighborhoods: An Analysis of Average and Contextual Effects," *Social Forces*, 84(3): 1697-1724.

Soltas, E. (2021): "The Price of Inclusion: Evidence from Housing Developer Behavior," forthcoming, *Review of Economics and Statistics*.

Su, Y. (2022): "The Rising Value of Time and the Origin of Urban Gentrification," *American Economic Journal: Economic Policy*, 14(1): 402-439.

Thomas, T., C. Hartmann, A. Driscoll, K. Chapple, A. Cash, R.R. Elias, & M. Zuk (2020): "The Urban Displacement Replication Project," UC Berkeley, https://www.urbandisplacement.org/wp-content/uploads/2021/08/udp_replication_project_methodology_10.16.2020-converted.pdf.

Towe, C. & C. Lawley (2013): "The Contagion Effect of Neighboring Foreclosures," *American Economic Journal: Economic Policy*, 5(2): 313-335.

Waldfogel, J. (2008): "The Median Voter and The Median Consumer: Local Private Goods and Population Composition," *Journal of Urban Economics*, 63: 567-582.

Whitaker, S. & T. J. Fitzpatrick IV (2013): "Deconstructing Distressed-property Spillovers: The Effects of Vacant, Tax-delinquent, and Foreclosed Properties in Housing Submarkets," *Journal of Housing Economics*, 22(2): 79-91.

Wong, F. (2020): "Mad as Hell: Property Taxes and Financial Distress," *mimeo*, UC Berkeley.

Zhang, D.H. (2022): "Closing Costs, Refinancing, and Inefficiencies in the Mortgage Market," *mimeo*, Harvard.

Online Appendix to

Property Tax Sales, Private Capital, and Gentrification in the U.S.

by Cameron LaPoint (Yale SOM)

Contents

A Legal Background on Tax Sales **48**

A.1 General Tax Sale Procedures . . . 48

A.2 Auction Procedures by State . . . 49

A.3 Differences between Mortgage and Tax Lien Foreclosure . . . 52

B Details on Mapping Gentrification **55**

B.1 Two-Stage Sorting Procedure . . . 55

B.2 Crosswalking Census Tracts . . . 56

B.3 Alternative Crosswalk from Aggregating Block Groups . . . 58

B.4 Additional Two-Stage Model Results . . . 59

C Additional Ring Analysis Results **64**

C.1 Baseline Results Using Differences in Ring Means . . . 64

C.2 Additional Foreclosure Wave Regression Results . . . 68

C.3 Empirical Derivatives Procedures . . . 69

C.4 Additional Results Using Empirical Derivatives . . . 70

D Pricing Foreclosure Options: The Repeat Distress Index **74**

E Sample Local Tax Delinquency Documents **81**

F Linking Private Equity Deals to Tax Sale Records **86**

List of Figures

A.1 Tax Sale Bidding Method by State . . . 50

A.2 Distribution of Lender's Money on the Table . . . 54

B.2 Tract Type Proportions by State (2005–2019) . . . 62

B.3 County-level Mapping of Nationwide Gentrification . . . 63

C.3 Pooled Inner Ring Estimates by Distance to Tax Sale . . . 67

C.6 Estimates of Tax Sale Pricing Spillovers (Non-institutional Lien Buyers) 75
D.1 Repeat Distress Index for Tax Liens: Total Bid Value 79
D.2 Repeat Distress Index for Tax Liens: Surplus Value 80
E.1 Example: Buyer's Book Page from July 2018 Tax Sale 83
E.2 Example: Final Notice Cover Page for Delinquent Taxpayer 84
E.3 Example: Final Notice Payment Stub for Delinquent Taxpayer 85
E.4 Example: Buyer's Book Total Proceeds from July 2018 Tax Auction 85

A Legal Background on Tax Sales

In this appendix section, I describe in detail methods for conducting tax sales – how they proceed in general and state-specific legal variations which impact investors' returns. I produce summary statistics which demonstrate why tax foreclosure is a (nearly) disjoint event from mortgage foreclosure. The information from this section originates from consulting the relevant state and local property tax statutes and practitioners' guides to tax lien investing (e.g. Loftis 2007).

A.1 General Tax Sale Procedures

On top of the three broad tax sale types described in Section 2.1, there are other tax sale policy parameters which vary across locality and influence investors' returns and ability to take possession of the property. Before delving into how bidding procedures can differ, I start by describing common conditions under which tax sales occur:

(i) Once the local tax authority sets an auction date, the properties up for auction are posted on a website for investors to review – typically at least two weeks in advance of the auction date. Properties can still be redeemed once listed on a tax sale advertisement, and edits to the sale notice list are common, which as discussed in Section 4.4 renders last-minute redeemed properties and their neighborhoods a natural control group for assessing local spillovers.

(ii) The vast majority of tax sale auctions are still conducted in person, and some states conduct annual tax sales with dates determined by statute.[1] As of writing, localities in Alabama, Arizona, California, Colorado, Georgia, Florida, Indiana, Mississippi, Tennessee, Texas, and Washington conduct sales online. For online sales, auctions proceed with sealed prices and lots sold at a very quick pace – sometimes with as little as 30 seconds in between properties – but the bid units differ by state (see Figure A.1 below).[2]

(iii) To register as a bidder, one must provide an SSN or EIN. This is a requirement at sign–up, because a W-9 must be filed for any potential income attached to the transaction. In some cases, bidders can sign up through an institution or retirement account rather than as an individual. In the event that an asset (either a lien or property) is purchased, the SSN or EIN of the bidder determines the lien or deed ownership, and the name attached to the identification number is the one which appears in tax sale records. For example, if an investor uses the SSN of their family member to place bids, then the asset will automatically be in the family member's name unless ownership is subsequently transferred.

(iv) Some counties may impose minimum fees that investors must place with the clerk in order to bid. Larger counties where properties are more valuable require a security deposit, which either gets refunded if the investor buys assets below that amount, or becomes automatically

[1] In-person auctions were suspended during the initial COVID-19 outbreaks, so tax sale data are largely missing for 2020 and parts of 2021. In counties which suspended their annual tax sale, many properties which would have been up for sale in 2020 were listed in 2021 or 2022.

[2] See the guidance for prospective bidders on Govease, which is a common platform used by governments to implement tax auctions online: `https://govease.helpscoutdocs.com/article/136-getting-started-with-govease`.

deducted from the invoice on any purchases.[3] For online auctions, the deposit must be remitted through a mailed check or electronic order to the tax authority in advance of the auction. Funds must also be verified, which can be more or less restrictive depending on the jurisdiction.

A.2 Auction Procedures by State

States differ in how bidders quote their bids, and how competing bids are resolved in favor of one party's ownership of the asset. There are five possible bidding types set out by state statutes. Figure A.1 maps each state to its predominant bidding type.

- **Premium bid:** By far the most common method for submitting bids (39 out of 51 states) is to quote a premium in relation to the outstanding tax bill. In that sense, this is a standard English price auction. There are subtle variations on this bidding method, where states may differ on the starting bid and/or place a floor on the minimum premium bid. The most common starting bid is the outstanding tax bill, so that the tax authority is always made whole if the asset is sold.[4]

- **Bid down interest:** This is a Dutch (descending-bid) auction where the bids are in terms of the interest rate that a buyer is willing to accept on a lien. By definition, a bid down interest auction only occurs in states engaged in lien sales. The starting bid is a maximum guaranteed annual interest rate which typically matches the interest rate levied by the tax authority on debt accumulated prior to tax sale. Successful bidders pay the amount of the outstanding tax debt and then are entitled to the final bid interest rate throughout the redemption period.

- **Random/rotational bid:** The tax authority sets a "buy it now" price for each property. Bidder numbers are then either called at random for each lot, or called in rotation in a randomly selected but fixed order. When a bidder is called they simply respond whether they would like to purchase the asset, and the asset is sold to the first bidder who answers affirmatively.

- **Bid down price:** This method is the same as bid down interest, except the bids are in terms of prices instead of interest rates. Currently, only Illinois and Louisiana use versions of this bidding method (both lien states).

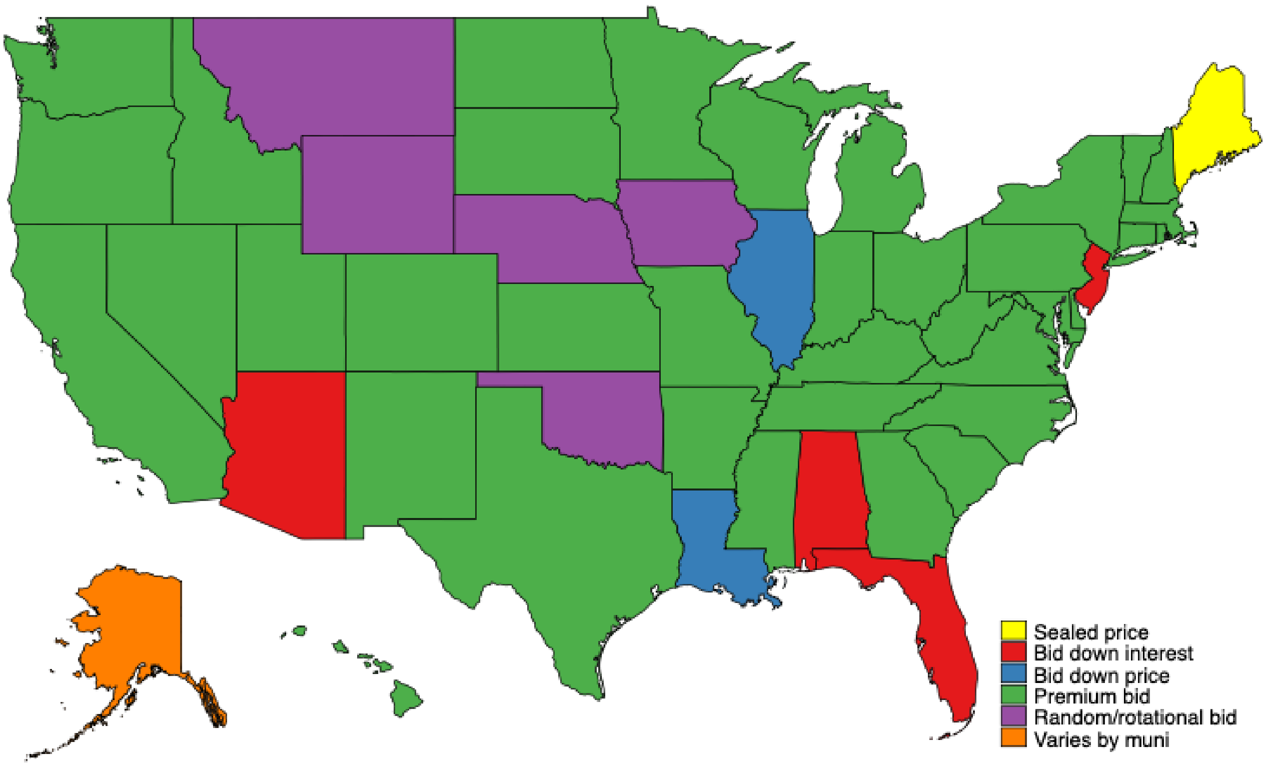
- **Sealed first price (Vickrey):** Currently, only Maine conducts sealed first price auctions.

Three other parameters are key to determining investor returns to trading tax sale claims: interest rates on tax debt, redemption period, and maturity in the case of liens. For lien states, redemption periods can range from 6 months (D.C. and Maryland) to 3 years (many states),

[3]Counties differ on whether the deposit must cover the full bid amount before a sale can be considered valid. This means that, strategically, institutional investors who are very active in a tax sale market will place a large deposit to avoid the risk that a property is called up for auction at a point when their total bids have exceeded the deposit amount. In such cases where bidders have already pledged their deposit limit on auction purchases, bidders may need to leave the auction room and replenish their deposit account before continuing to bid. In D.C., there is a $200 flat administrative fee an investor pays if they make a purchase, and a required deposit amount of 20% of the total eventual bids.

[4]Another key variation within English auctions is which component of the lien amount can accrue interest. For example, in D.C. only the lien amount can accrue interest, so any premium in excess of the tax bill starting bid does not generate separate interest income for the lienholder.

FIGURE A.1. Tax Sale Bidding Method by State



Notes: The map classifies states into six distinct categories based on how investors at tax sale auctions submit bids for claims to tax-delinquent properties under state law. Premium bid refers to an auction method where the starting bid is set so that the tax authority breaks even (i.e. the bid is equal to the outstanding tax debt plus any interest and penalties), and investors may make bids on top of this premium. In bid down interest states, prospective lienholders compete by bidding the lowest interest rate on the lien they are willing to accept, with the starting bid set at a statutory maximum annual rate; bidding continues until an investor offers 0% interest, in which case the bid is set to the premium – as in a premium bid auction. Auctions in bid down price states operate similarly to those in bid down interest states, except the bid is the lowest value (Illinois) or ownership share (Louisiana) the investor is willing to accept. Random or rotational bidding states select auction card numbers at random and ask the bidder whether they want to purchase the claim, where the take it or leave it offer is set so the tax authority breaks even and the lienholder receives a fixed statutory interest rate. Only Maine engages in sealed (first) price auctions. Alaska is the only state in which the auction bid method is statutorily allowed to vary by municipal rule. See text for details on differences in how local governments may conduct auction proceedings and implications for investor yields. Information sourced by the author directly from the relevant statutes regarding property tax collections in each state.

and the expiration date can range from 6 months (D.C.) to 20 years (New Jersey) after the lien purchase date. 12 of the 23 deed states have no redemption period, wherein a tax sale results in a final loss of title. In most lien states, the redemption period and expiration or maturity period are the same, meaning that at the end of the redemption period, the lienholder must decide whether to foreclose and cannot continue collecting interest/penalties on the tax debt.

Example: calculating returns to a tax lien portfolio. To illustrate how the tax sale policies set by each state translate to returns, consider an investor who purchases the following bundle of interest-bearing lien certificates tied to properties in Washington, D.C., Florida, and Massachusetts.

Sample Portfolio of Tax Lien Certificates

	Washington, D.C.	Florida	Massachusetts
Sale method	Lien	Lien	Hybrid
Bid method	Premium bid	Bid down interest	Premium bid
Redemption period	6 months	2 years	6 months
Maturity/expiration	6 months	7 years	6 months
One-time penalty rate	0%	5% minimum	0%
Annual interest rate	19.5%	18%	16%
Assumed total bid value	$16,000	$1,000	$100,000
Assumed premium bid	$2,000	N/A	$75,000

Notes: For each lien in the table I assume a bid value (i.e. the face value) equal to the median I observe in my data; this is the actual amount (exclusive of administrative fees) initially paid by the investor to obtain the property lien. In Florida where the interest rate is determined by final auction bid, the modal interest rate on liens in my data is the statutory maximum 18%, indicating uncontested bids. All other parameters in the table are set by statute.

The liens carry different interest rates. In D.C. and Massachusetts, which conduct premium bid auctions, the interest rate is set by statute. In Florida, the interest rate is the lowest rate bid at auction, where bidding starts at 18%; hence, we know from this example that the investor obtained the Florida lien via an uncontested bid. To shield investors against the risk of quick redemption, Florida offers a minimum penalty of 5% on the face value of certificate if the total interest amount earned is less than this threshold. Since D.C. is a pure lien state, the interest only accrues on the outstanding tax debt, not on the premium bid. In contrast, in Massachusetts the interest also accrues on the premium portion of the total bid because the investor receives a hybrid deed The higher bid value for Massachusetts reflects the fact that unless the delinquent taxpayer redeems the lien, the investor automatically holds a deed to the property (see Section 2.1 for more details).

With this information in hand, it is straightforward to compute the yields on the portfolio.[5]

[5] Generally once a new month of the redemption period begins, the interest automatically accrues for that entire month. For simplicity, I assume all liens were acquired as of the first of the month in this example, meaning I implicitly adopt this convention for interest accrual on lien certificates. Some states are explicit on this in their property tax law, but others are not.

In this context, the lien is a bond that "matures" either (i) when the redemption period ends and foreclosure proceedings conclude, or (ii) when the certificate expires, regardless of whether the investor moves to foreclose. In the former scenario, the face value of the bond gets rolled into the price paid for the property at foreclosure (i.e. the "credit bid" at foreclosure). In D.C. and Massachusetts the redemption period length is the same as the maturity, and in Florida the maturity is 7 years compared to a 2-year redemption period. Investors moving to foreclose off of a lien is relatively rare in FL given the long redemption period and steady stream of interest payments accruing at high annual rates, and consequently, the face value is quite low. Generally, the divergence across states in face value of the liens reflects how the auction regime impacts the probability, or desirability, of entering into foreclosure.

Suppose the investor with this portfolio holds each lien until the end of the redemption period, but does not exercise the foreclosure option at the end of the redemption period in D.C. and Florida. By default, in Massachusetts, this investment strategy would entail possessing a deed (but not title) to the property at the end of six months. Equivalent to this strategy is the scenario where the delinquent taxpayer redeems the lien on the last day of the redemption period for each property. For D.C. the monthly interest rate is 1.5%, so at the end of the six-month redemption period the interest payment received is $(\$16,000-\$2,000)\times[(1+0.015)^6-1] = \$1,308.21$. The investor realizes a $1,308.21/$16,000 = 8.2% yield, since the interest only accrues on the tax debt portion of the bid. Turning to the Florida lien, the total interest at the end of two years is $\$1,000\times[(1+0.18)^2-1] = \$392.40$ at the annual 18% rate. The minimum guaranteed coupon is the 5% penalty, which in this case would be the lower amount of $50, so the annual yield is the 18% interest rate bid. Finally, in Massachusetts, where the implied monthly interest rate is 1.24%, the taxpayer must redeem the full credit bid plus interest of $\$100,000\times[(1+0.0124)^6-1] = \$7,703.30$, even though the initial value of the tax debt was only $25,000. Hence, the yield on the MA lien is 7.7%. The MA lien illustrates how in hybrid states, excessive premium bidding can create a substantial hurdle to lien redemption.

A.3 Differences between Mortgage and Tax Lien Foreclosure

How closely related are mortgage and tax foreclosures? A short answer is that the two types of foreclosure are almost entirely disjoint events. To establish this fact, I merge my sample of D.C. tax liens to CoreLogic Deeds using the square and address combination. This merge yields the full set of title exchanges associated with the property and details about any mortgage contracts and lenders involved in the transaction. Of the 15,865 liens merged to CoreLogic, 1,248 experience a title change *within* the six-month statutory redemption period, and of the latter sample, 146 experience a mortgage foreclosure. Thus, 146/1,248 = 11.7% of redeemed liens undergo mortgage foreclosure, but only 146/15,865 = 0.9% of all tax liens do. This is compared to a 4.1% foreclosure rate for all title exchanges in D.C. over the same sample period (2005-2019).

What explains the relatively small proportion of bank foreclosures on tax-distressed properties? There are two main reasons for these empirical patterns. First, on the taxpayer side, most lenders require an escrow account with monthly payments tacked onto the monthly loan payment. For taxpayers without a mortgage who lack an escrow account and are predominantly older homeowners, there is a higher chance that they forget to make tax payments and thus fall into delinquency. Owing to this, 54.8% of tax sale properties do not have an attached mortgage as of the last transaction in CoreLogic involving at least one current, tax-delinquent owner.

Second, for the remaining 45.2% tax sale properties last purchased with a mortgage, the capital the lender has remaining in the house is either minimal relative to the outstanding tax debt to be redeemed, or zero based on the loan's implied amortization schedule. 5.1% of tax sale properties have a loan term which ends before the tax lien sale date. Even in cases where the loan due date is unavailable, to conservatively estimate the balance I assume any remaining loans have a 30-year term.[6] Applying standard accounting formulas to find the remaining balance given the number of months since origination, another 9.3% of the sample has a zero loan balance.

More concretely, I conduct the following steps to estimate the remaining loan balance for properties where the mortgage term is either missing or ends after the tax lien sale date:

1. Identify the loan contract type, including the origination amount C_0, the quoted mortgage rate i, fixed rate mortgage (FRM) vs. adjustable rate mortgage (ARM), and the loan term in months N. All loans in my sample charge payments at a monthly frequency.

2. Drop any observations where there is record of a refinancing event or origination of a home equity loan. For such loans, the amortization schedule cannot be determined since the loan contract – notably the remaining monthly payments – will differ from the initial mortgage.

3. For FRMs, the following standard accounting formulas pin down the monthly payment PMT and the balance C_n after n months of payments:

$$C_0 = \sum_{t=1}^{N} \frac{PMT}{(1+i/12)^t} \implies PMT = C_0 \times \left(\frac{i/12}{1-(1/(1+i/12))^N} \right) \tag{A.1}$$

$$C_n = (1+i/12)^n \times C_0 - \sum_{t=0}^{n-1} (1+i/12)^t \times PMT$$

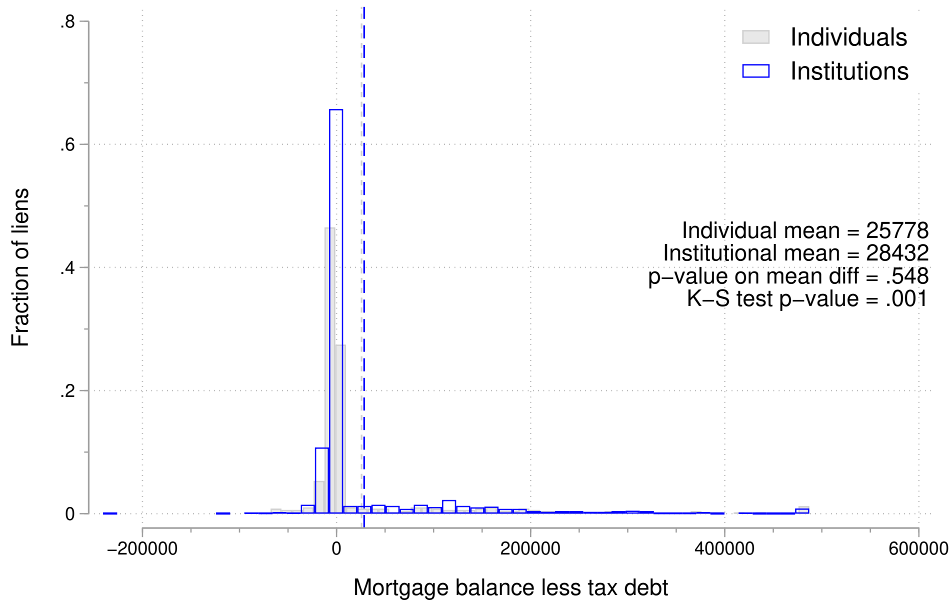
$$\implies C_n = (1+i/12)^n \times C_0 - \frac{PMT \cdot \Big((1+i/12)^n - 1 \Big)}{i/12} \tag{A.2}$$

   where (A.1) starts with the fact that the loan fully amortizes over N months, and expands the geometric series to obtain PMT. I derive equation (A.2) by iterating on the initial law of motion $C_1 = (1+i/12) \times C_0 - PMT$.

4. For ARMs, more assumptions are necessary to track the mortgage balance since the PMT will vary over time with fluctuations in the index rate. I assume standard ARM contract terms: 5/1 loan (i.e. 5-year initial period, followed by annual rate reset), a 1-year Treasury bill index plus constant 2.75% margin, and a 2/2/2 cap structure. Unfortunately, CoreLogic Deeds does not record these features. According to Freddie Mac Mortgage Market Survey data, lenders' margins on 5/1 ARMs have been virtually constant over the last 20 years. The resulting balances are virtually identical if I instead pretend each ARM is an FRM with a constant rate fixed at the initial rate. This is because while the PMT continually adjusts downward over the amortization schedules within my sample period, the fraction of PMT going towards principal will also adjust to allow the loan to fully amortize over the term. The accounting is again given by (A.1) and (A.2), except with the ARM at the beginning of each reset period n months into the loan, I set replace C_0 with C_n and then update i_t according to the index + margin and cap structure.

[6] That is, to the extent that the true loan term is less than 30 years, this assumption will result in an overestimate of the principal balance remaining as of the tax sale date.

FIGURE A.2. Distribution of Lender's Money on the Table



Notes: The figure plots the distribution of the gap between the lender's implied stake in a property and the redeemable tax debt attached to the local tax lien. The sample is D.C. tax lien sale properties which can be matched on the basis of the address and delinquent taxpayer to mortgage records in CoreLogic Deeds. I report the mean gap for the overall sample and the mean for the subset of liens purchased by individual vs. institutional investors. I winsorize the gap at the 99.5th percentile. See text for details on the calculation of loan balances, which follow the accounting identities in (A.1) and (A.2). I provide p-values for a test of the difference in means and Kolmogorov-Smirnov tests of the difference in the distributions between institutional and individual investors.

5. For all remaining loans where I cannot identify the term or the contract type but observe the initial quoted rate, I assume the loan is a 30-year FRM. This results in conservative overestimates of the remaining loan balance as of tax sale to the extent that the vast majority of loans, including ARMs, have terms of 30 years or less.

Implicitly, this algorithm assumes the borrower makes continuous monthly payments, without any breaks due to debt discharge or forbearance plans. The former would push balances towards zero, while the latter would mean I underestimate terminal balances.

Figure A.2 plots the distribution of the gap between the implied mortgage balance on the loan and the tax debt on the lien. A value of 0 would indicate the lender could, at best, break even by redeeming the tax lien. I say "at best" because this notion of lender money on the table does not take into account legal fees the lender might incur through foreclosure, fees charged by the tax authority for processing the redemption, or interest and penalties accruing to the lienholder within the redemption period. A negative value implies the tax debt is greater than the capital at stake, and so it is profitable for the lender to not redeem the lien. In over 80% of the cases, it would either be not possible or unprofitable for the lender to redeem. There is no difference, on average, between the money at stake for the lender in cases where an institution vs. individual is the lienholder, suggesting neither investor type is better at targeting properties without competing liens.

B Details on Mapping Gentrification

This appendix presents the full two-stage sorting model used to classify gentrifying vs. non-gentrifying areas, including the crosswalking methods for defining population flow variables across decennial Census geographies.

B.1 Two-Stage Sorting Procedure

I perform five major steps, listed below, to characterize neighborhoods. Although I estimate two versions of the model – one using data between 1990 and 2005 and another using data between 2005 and 2019 – I write the steps in terms of two general endpoints t and $t+1$, where t and $t+1$ could potentially lie within different decennial Census geographies T and T', necessitating the need to crosswalk tract boundaries between T and T'. I also write the algorithm in terms of the three general "tipping point" parameters $\{x_1, x_2, x_3\}$.

1. **1st Stage.** I create growth measures, $\Delta X_j = X_{j,t+1} - X_{j,t}$, where $X_{j,t}$ is a Census variable pertaining to tract j in time t. Tract j is defined according to the latest geography T'. The model relies on three growth measures:

   (i) The change in the overall number of non-low-income individuals.

   (ii) The change in the overall number of low-income individuals.

   (iii) The change in the population *share* of low-income individuals.

   Low-income individuals are those below 200% of the poverty line. I use population counts by ratio of income to the poverty line which are defined at the *individual level* to construct these variables. The results are materially unchanged if instead I count families or households, but individual population accounts have the advantage of being available at fine geographic levels going back further in time. Because the definition of the federal poverty line changes over time with inflation no additional adjustment is needed for changes in living standards.

2. I generate a flag, *expanding*, indicating whether the T' Census tract j is economically thriving based on in-migration of high-income individuals. The standard spatial sorting condition for labor à la Rosen-Roback rationalizes this definition. A tract is expanding if both of the following are true:

   (a) The absolute number of non-low-income individuals increased by more than x_1% between t and $t+1$.

   (b) The population share of low-income individuals declined by more than x_2 p.p. between t and $t+1$.

   Similarly, a tract is *declining* if both of the following are true:

   (a) The absolute number of non-low-income individuals declined by more than x_1% between t and $t+1$.

   (b) The population share of low-income individuals increased by more than x_2 p.p. between t and $t+1$.

3. **2nd Stage.** Create a dummy for low-income population growth. Using the 1st and 2nd stage definitions, define *TractType*

- $= 1$ if economically expanding + low-income population growth $\geq x_3\%$. I refer to such areas as a *growth* tract.
- $= 2$ if economically declining + low-income population growth $\geq x_3\%$. I call these areas *low-income (LI) concentration* tracts.
- $= 3$ if economically expanding + low-income population decline $\geq x_3\%$. This is my measure of a "low-income crowd-out" or *gentrification* tract throughout the paper.
- $= 4$ if economically declining + low-income population decline $\geq x_3\%$. Consistent with anecdotes and predictions of structural models (Owens, Rossi-Hansberg, Sarte 2020), a large fraction (40.1%) of Census tracts in central Detroit relative to other metros fall into the *abandonment* tract category, with a majority (45.5%) classified as *LI concentration*, 14.4% classified as *gentrifying*, and only 1.0% are *growth* tracts.
- I code *TractType* as *unclassified* if the tract does not fulfill the criteria for any of these four categories. In such cases, the tract did not experience substantial turnover in the income composition of its residents, conditional on the parameters $\{x_1, x_2, x_3\}$ set for the period between t and $t + 1$.

4. **Weak vs. Strong Type Categories.** The above steps result in a 4-type classification of Census tracts. As a refinement, I expand to an 8-type classification by introducing a "weak" version of each of the four values for $TractType$. As a baseline, I use $\{\overline{x_1}, \overline{x_2}, \overline{x_3}\} = \{10\%, 5\text{p.p.}, 0\%\}$ as the cutoff vector for strong tract type definitions, and $\{\underline{x_1}, \underline{x_2}, \underline{x_3}\} = \{5\%, 2.5\text{p.p.}, 0\%\}$ as the set of cutoffs for weak tract type definitions. Any tracts not satisfying the looser set of cutoffs remain in the *unclassified* category, and any tracts with $\underline{x_1} \leq x_1 \leq \overline{x_1}$ and $\underline{x_2} \leq x_2 \leq \overline{x_2}$ are classified as the weak version of $TractType$.

5. **Peristence.** Finally, while I estimate the model separately for two subperiods $t = 1990$ and $t + 1 = 2005$, or $t = 2005$ and $t + 1 = 2019$, I assess the stability of the tract type definitions across subperiods by fixing boundaries as of $T' = 2010$ That is, for the 1990-2005 subperiod, I crosswalk the 1990 tracts to 2000 tracts, and then transform the variables again to render them consistent with 2010 geography. The end result is a set of transition probability matrices of tract types. From this matrix I focus on two summary measures of persistence:

   (I) *Type persistence*: This is a dummy equal to unity if tract j has the same type – either in the 4-type or 8-type version of the model – in each subperiod.

   (II) *Gentrification persistence*: In each subperiod, I create a flag equal to one if the tract is either weak or strong gentrifying, and equal to zero otherwise. The correlation of this dummy between the two subperiods and across Census tracts indicates the joint likelihood that either a gentrifying area continues to gentrify or that a non-gentrifying area will gentrify in the future. If this correlation is closer to zero, it indicates neighborhoods are in a steady state; if it is closer to one, then it indicates gentrification persists over 30 years. Finally, as this correlation moves more negative, the more neighborhoods are subject to income-based tipping points over 15-year intervals.

B.2 Crosswalking Census Tracts

This method follows the procedures of the Opportunity Atlas constructed by Chetty et al. (2018), among others. The idea is to rescale the population variables under the older Census boundaries by the fraction of the old tract population (i.e. based on T) that lives within the new tract boundaries (i.e. based on T'). I use mapping between 2000 and 2010 as my main example, but a similar exercise can be done for other decennial Census geographies, and I discuss below how the steps vary for converting between 1990 and 2000 Census tract boundaries. The steps are as follows:

1. Download 2010 Census Tract Relationship files (for all 50 states + DC) from the U.S. Census Bureau. There are three categories of files: the relationship files, and lists of substantially changed 2000 Census tracts and 2010 Census tracts; these latter two files can be used to create a crosswalk which only accounts for tract changes where a tract split into multiple tracts. I account for all Census tract changes, which are reflected in the relationship files.

2. Conceptually, take a 2000 Census tract j which is now (partially) contained in 2010 Census tract k. Each tract is a collection of blocks (the smallest geographic unit) or block groups, indexed as $j_1, j_2, \ldots j_n$ and $k_1, k_2, \ldots k_m$. In cases where there is one entry for a 2000 tract code (TRACT00), then the tract did not substantially change. When there are multiple entries for a 2000 tract code, then the tract split into several tracts after the 2010 Census.

3. Consider a variable (e.g. total population) from the pre-2010 Census year t for 2000 tract j: $X_{j,t}$. We define $X_{k,t}$ at the 2010 Census tract level via:

$$X_{k,t} = \sum_{j \subseteq k} \omega_j \cdot X_{j,t} \tag{B.1}$$

   where ω_j is a weight that accounts for the fraction of the 2000 tract population that overlaps with the new Census tract k. Start by setting ω_j = POPPCT00 from the relationship files, which in words is defined as:

$$\text{POPPCT00} = \frac{\text{2010 population of the overlap}}{\text{2010 population of the 2000 tract}} \tag{B.2}$$

   This effectively sets the 2010 Census tract value of the variable in t equal to the population-weighted average of the variable across all 2000 Census tracts which now overlap with the new tract. In cases where j has complete spatial overlap with k, mechanically $\omega_j = 1$. Importantly, even if j is completely contained within tract k, there will be many cases where $\omega_j < 1$ because the new tract absorbed pieces of other old Census tracts.

4. To create the model inputs, I compute changes in population variables using the newer tract definition. More generally, $\Delta X_k = X_{k,t+1} - X_{k,t}$, where k is defined according to the latest geography T', and $t+1$ and t are the endpoints over which I estimate the model (e.g. t = 2005 ACS and $t+1$ = 2019 ACS).

5. Finally, to compare versions of the model estimated using data from 1990-2005 vs. 2005-2019, I crosswalk tracts under the 1990 Census geography to 2010 boundaries. This involves first crosswalking a 1990 tract i to a 2000 tract j, and then applying steps 1 through 3 above to create the $X_{k,t}$ where tract k is a 2010 Census tract. That is, I first compute:

$$X_{j,t} = \sum_{i \subseteq j} \omega_i \cdot X_{i,t} \tag{B.3}$$

$$\text{where } \omega_i = \text{POPPCT90} = \frac{\text{2000 population of the overlap}}{\text{2000 population of the 1990 tract}} \tag{B.4}$$

and then feed in the resulting $X_{j,t}$ into equation (B.1). To calculate (B.3), I use the 1990 to 2000 Census population-based relationship files. Note the same exercise as in all of the above steps can be performed with county as the geographic unit.

B.3 Alternative Crosswalk from Aggregating Block Groups

An alternative way to crosswalk tract boundaries between decennial Censuses would be to define a new variable POPPCTOLD which uses 2000 population rather than 2010 population to define this weight:

$$\text{POPPCTOLD} = \frac{\text{2000 population of the overlap}}{\text{2000 population of the 2000 tract}} \tag{B.5}$$

Why does this potentially matter for computing long-run growth rates? Consider a simple example. 2000 Census tract 1000 gets split across three new tracts (1001, 1002, 1003) as a result of the 2010 Census. In the 2000 Census, the portion of 1000 that now corresponds to 1002 had a population of 5,000 people, compared to a total population of 10,000 for tract 1000. This means $\omega_j = 0.5$ under the definition of POPPCTOLD.

However, suppose there was a chemical spill that occurred in the 1002 tract area between the two decennial Censuses, so that now the 2010 population of the overlap is only 1,000, compared to a 2010 population of the 2000 tract boundaries of 10,000. This would imply a weight of $\omega_j = 0.1$. If we try to compute growth rates for a variable $X_{k,2000,t}$ between 2000 and some post-2010 year t, then using POPPCT00 as the weight is going to lead to a much larger growth rate, because the weight takes into account the fact that people fled the chemical spill area. Which method generates a more accurate depiction of population movements between 2000 and post-2010 years depends on whether the chemical spill occurred closer to 2000 or 2010. If the incident occurred closer to 2000, then using POPPCTOLD is more accurate. However, if the event occurred in, for example, 2009, then $X_{k,2000}$ is severely underestimated.

The following steps describe how to construct POPPCTOLD by aggregating up blocks using the Census Block Relationship files:

1. NHGIS (IPUMS) provides a crosswalk between 2000 block groups and 2010 Census tracts that I use to calculate the numerator term in the weight POPPCTOLD (B.3). There are 52 crosswalk files – one for each state (+ D.C.) and a national file for the whole country. The crosswalk in these files is structured as a long character concatenation of several geographic subcodes, called a "GISJOIN" because it can be imported into the spatial software platform GIS. This character string for each year (2000 or 2010) includes the state, county, Census tract, and block group codes. To construct the numerator, I find the set of 2000 block groups that overlap with the same 2010 Census tract. To do this, I extract the portion of the 2010 GISJOIN that corresponds to the tract level (i.e. chop off the block group portion and select the county-tract code).

2. Next, I sum population counts across all the 2000 block groups that overlap with the 2010 tracts identified in the previous step. Note that the NHGIS crosswalk files include *wt_pop* which is the fraction of the source area's population located in the target zone; for example, the fraction of a 2000 block group's population located in the 2010 Census tract. The numerator

term in (B.5) can therefore be computed as:

$$\text{2000 population of the overlap} = \sum_{g \in k} wt\_pop_g \cdot X_{g,2000} \tag{B.6}$$

where $X_{g,2000}$ is the 2000 block group population, which can be found in the 2000 Census block group extract. The sum is across all 2000 block groups g attached to the same 2010 Census tract k.

3. Finally, I create an alternative measure of equation (B.1) by combining equations (B.5) and (B.6) to construct the alternative weights POPPCTOLD. The weights POPPCTOLD can then be substituted as the ω_j in equation (B.1) which translates variables, such as population by income to poverty ratio, from a 2000 Census tract j to the new 2010 Census tract k.

B.4 Additional Two-Stage Model Results

Here I summarize additional results from the two-stage model of population flows.

Figure B.1 explores how the proportion of tracts assigned to each of the four types changes as the threshold parameters become stricter. Each panel varies one of the three parameters x_1, x_2, x_3, holding fixed the other two. The dashed vertical line in each panel indicates the parameter adopted under the baseline "strong" version of the model. The first parameter governs whether a tract is expanding or declining based on absolute population flows, and thus is a way of classifying whether there is any significant income-based sorting. Being either expanding or declining in the first stage is a necessary condition for any of the four tract type definitions. Hence, all four tract type fractions are strictly decreasing in x_1. x_2 also controls whether a tract is expanding or declining but based on changes in population *shares* of low vs. high-income residents, and consequently all four tract types are strictly decreasing in x_2 as well.

The final parameter x_3 controls how severe low-income population growth must be in the second stage of the model for a tract to be assigned to one of the four types – otherwise the tract remains "unclassified," or not undergoing significant demographic shifts. The greater x_3, the higher low-income population growth must be for a tract to be classified as either LI concentration or growth. Conversely, since gentrification and abandonment tracts are defined by the strength of low-income population *decline*, increasing x_3 tips the scale in favor of each of those types. Of the four tract types, LI concentration is the most sensitive to the choice of cutoffs, especially in the first stage sort.

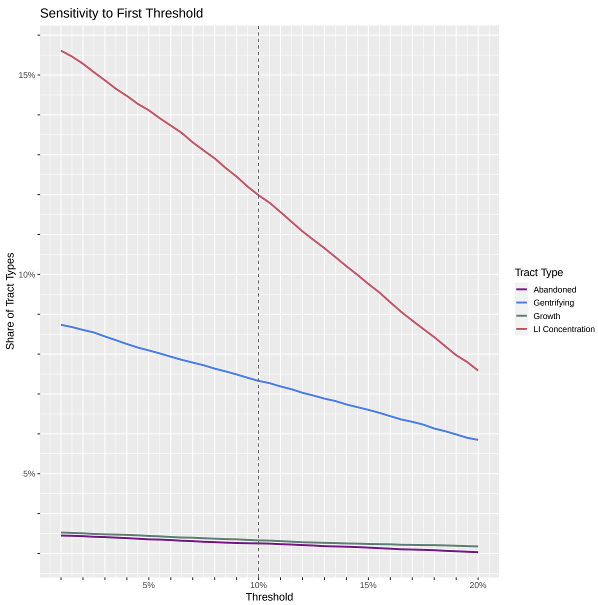
Figure B.2 reports the incidence of each tract type by state, replicating the analysis of Figure 4 for the later subperiod 2005-2019. Both gentrification and low-income concentration are noticeably more prevalent in more recent times. The model thus captures the trends towards increasingly entrenched regional income inequality documented elsewhere (e.g. Gaubert et al. 2021). The model also reflects more region-specific outcomes such as the TFP boom in North Dakota and Montana around 2007 due to shale gas exploration documented in Caliendo et al. (2018). This is evidenced by the large uptick of 15 p.p. in the share of gentrifying tracts in those states between the earlier and later subperiod.

Figure B.3 displays the spatial distribution of gentrification when I apply the two-stage model to all U.S. counties over the later subperiod (2005–2019). Panel A maps the county types for the strong version of the model with cutoff parameters $\{\overline{x_1}, \overline{x_2}, \overline{x_3}\} = \{10\%, 5\text{p.p.}, 0\%\}$, while Panel B does the same but for the weak version of the model with lower cutoff parameters $\{\underline{x_1}, \underline{x_2}, \underline{x_3}\} =$

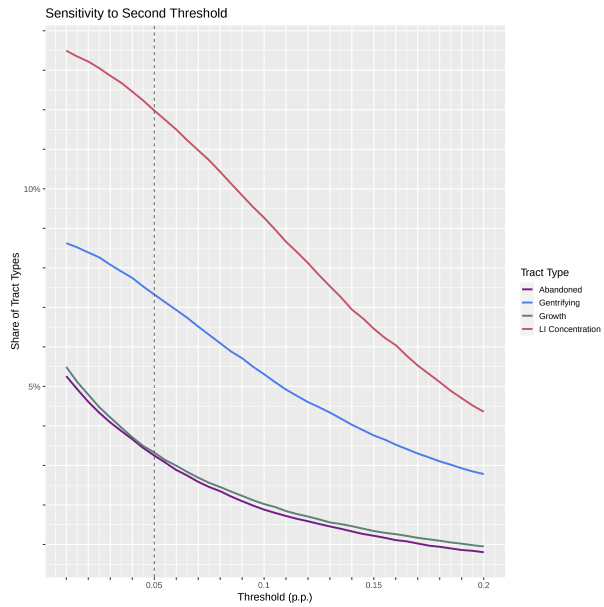
$\{5\%, 2.5\text{p.p.}, 0\%\}$. Both the strong and weak versions of the model assign LI concentration status to large swathes of the Rust Belt, including Ohio, Michigan, Indiana, and Illinois. The model picks up the recent economic expansions in the Southwest and Mountain states, including large areas of southwestern Texas, Montana, Arizona, and areas Idaho around Boise; many of the counties in those states are assigned gentrification and growth status. Abandonment is uncommon but occurs in rural pockets of the Deep South and severely distressed sections of the Rust Belt.

FIGURE B.1. Tract Type Proportions by Tipping Point Threshold Parameters

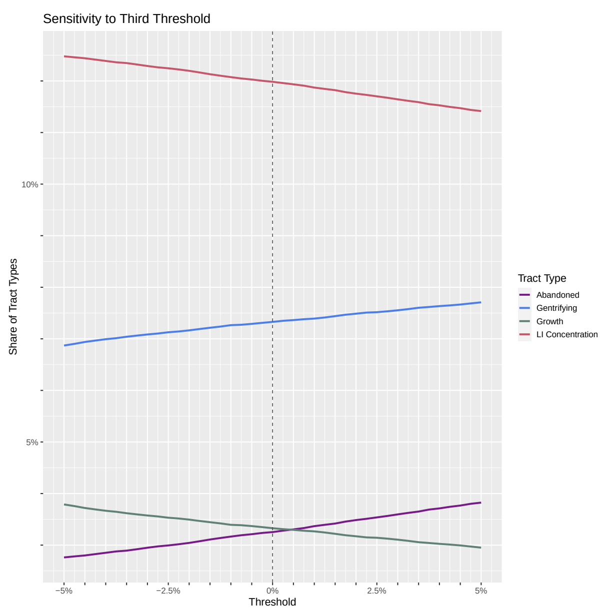
$\overline{x_1}$: 1st Stage Growth in Low-income Residents



$\overline{x_2}$: 1st Stage Change Low-income Share

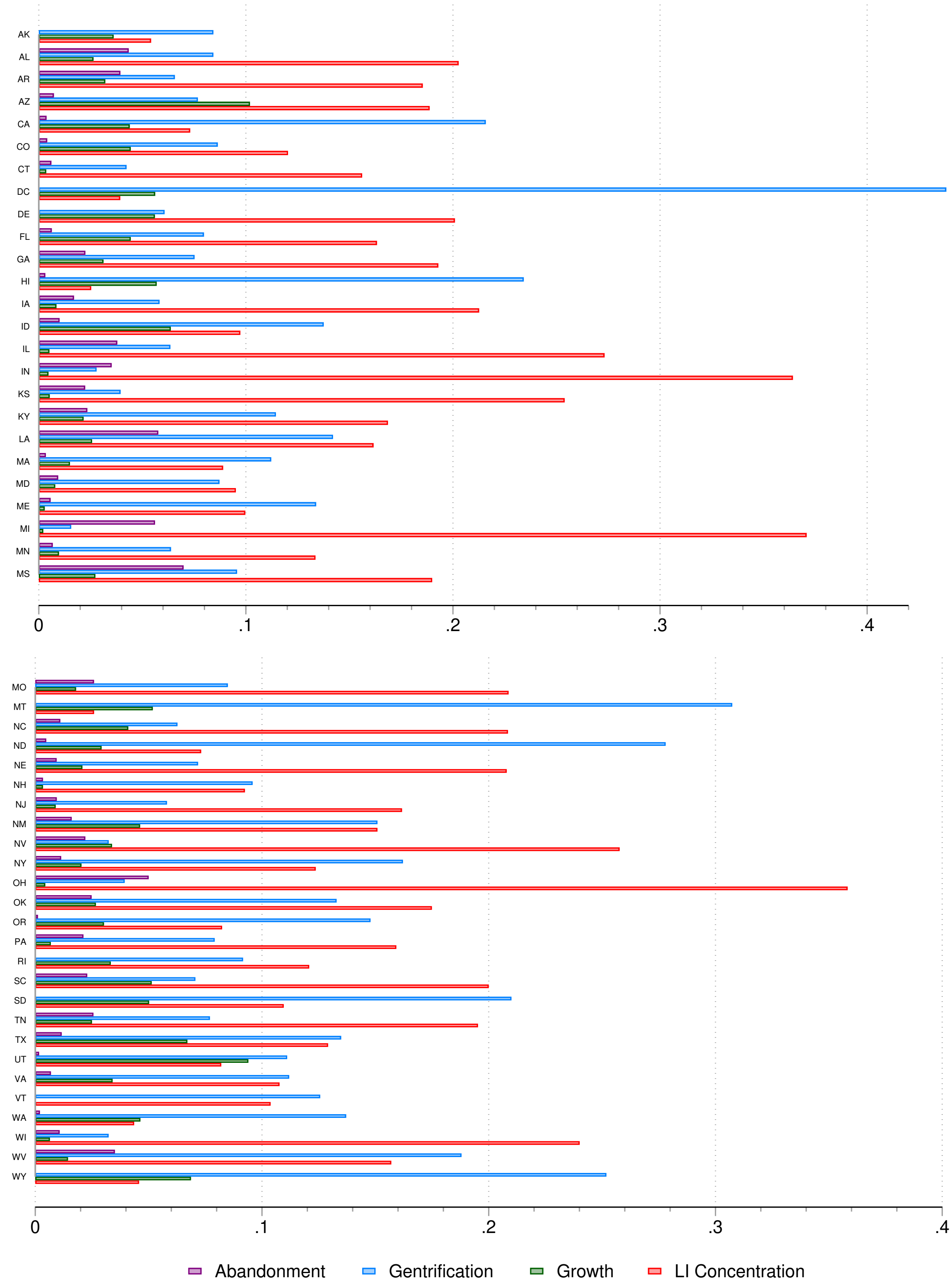


$\overline{x_3}$: 2nd Stage Growth in Low-income Residents



Notes: The figure displays how the tract type proportions in the 4-type version of the population flows model presented in Section 3 vary with respect to each of the three threshold parameters $\{\overline{x_1}, \overline{x_2}, \overline{x_3}\}$. In varying one parameter, I hold fixed the two other parameters. In each panel, the dashed vertical line indicates the baseline parameter value I use to estimate the model. The residual proportion of plots after accounting for the four types admitted by the model (not pictured) falls into the *unclassified* category. I conducted this sensitivity analysis for the 1990 – 2005 subperiod under the 2010 Census geography.

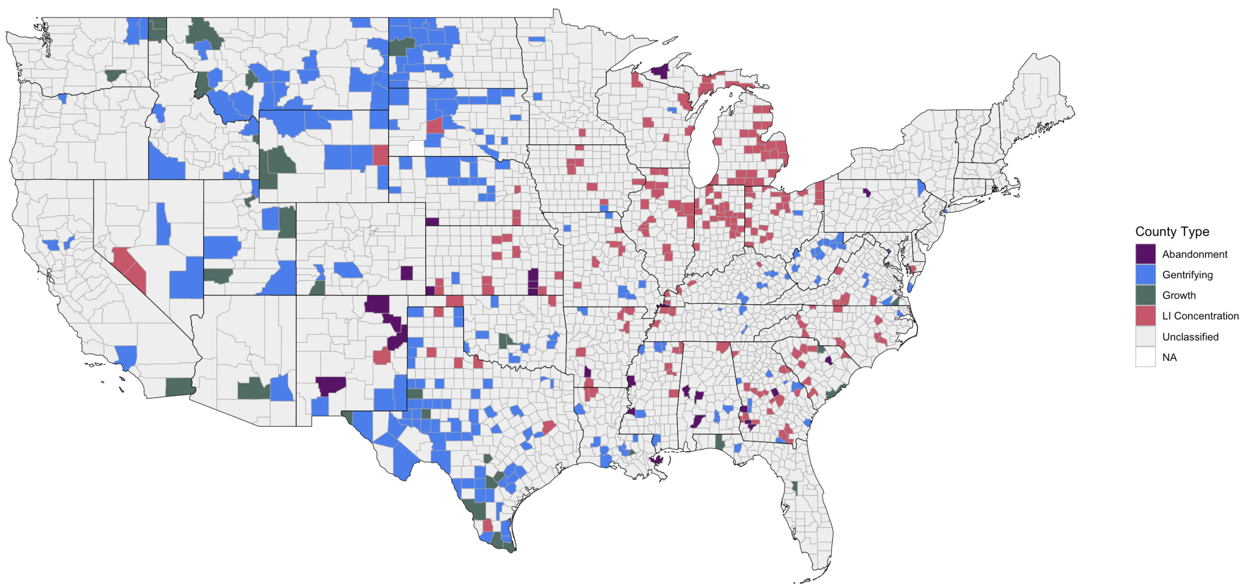
FIGURE B.2. Tract Type Proportions by State (2005–2019)



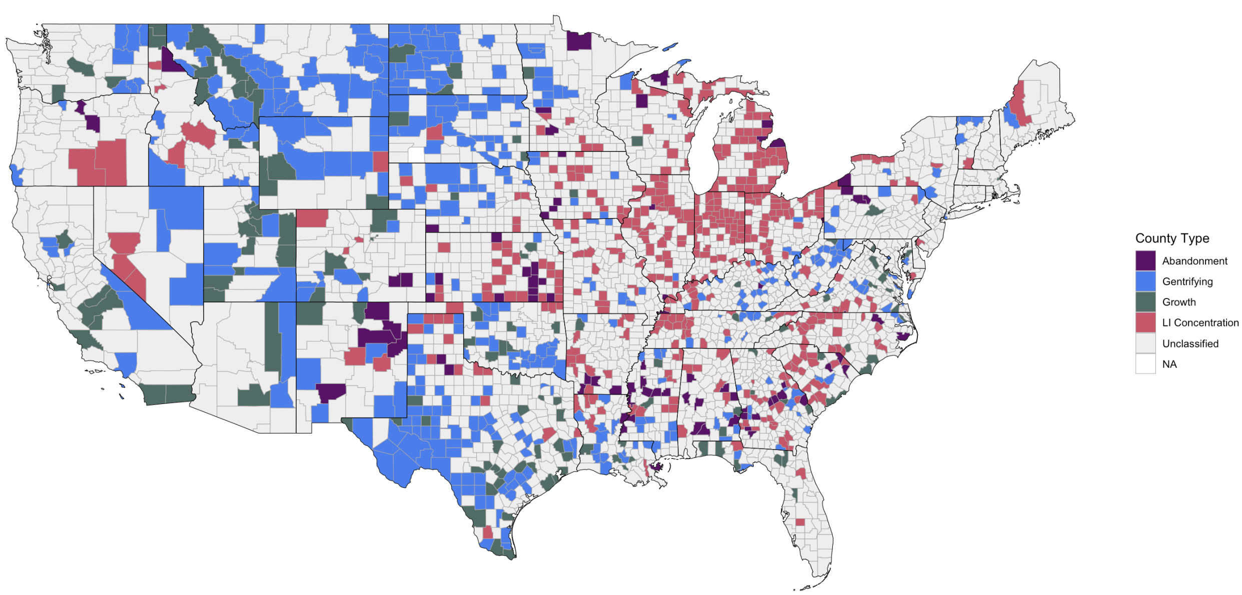
Notes: The figure shows the proportion of 2010 Census tracts in each state that fall into the four categories obtained from the two-stage classification model estimated over the period 2005-2019. Within each state, the tract type proportions do not sum to one, because the residual fraction of tracts fall into an *unclassified* category, which means the tract did not undergo major changes in income-based population sorting during the time period. I downloaded the 2005-2009 and 2015-2019 5-year ACS tract-level data from IPUMS NHGIS. See Appendix B.2 for details on the crosswalking methods.

FIGURE B.3. County-level Mapping of Nationwide Gentrification

A. 4-type Version of the Model ($\overline{x_1}, \overline{x_2}, \overline{x_3}$)



B. 8-type Version of the Model ($\underline{x_1}, \underline{x_2}, \underline{x_3}$)



Notes: The figure displays two maps classifying U.S. counties over the period 2005-2019 according to either the 4-type model with the "strong" set of parameters (Panel A), or the 8-type model with the "weak" set of parameters (Panel B). For readability, in the 8-type model of Panel B, I pool the weak and strong categories for each of the four major tract types. See Appendix B.1 for details.

C Additional Ring Analysis Results

In this appendix, I describe additional spatial difference-in-differences results obtained from taking simple differences in ring means, estimating versions of the foreclosure wave regressions in Section 4.1 directly comparable to those presented in preceding studies, and checking robustness of the empirical derivatives estimator to the choice of alternate tuning parameter sets.

C.1 Baseline Results Using Differences in Ring Means

A natural first step in spatial inference is a simple ring difference-in-differences (DD) design which compares the change in house prices in an inner or middle ring to the change in house prices in an outer ring.[7] I define the inner ring as the average house price transaction across properties i sold in year T within a 0.1 mile radius of a tax sale property location ℓ:

$$R^{in}_{\ell,T} = \frac{\sum_{i,t} p_{i,t} \cdot \mathbb{1}\{Yr(t) = T\} \times \mathbb{1}\{r \leq 0.1\}}{\sum_{i,t} \mathbb{1}\{Yr(t) = T\} \times \mathbb{1}\{r \leq 0.1\}} \tag{C.1}$$

where T denotes years relative to an institutional tax sale date + the end of any statutory redemption period (e.g. six months for D.C.), and $Yr(t)$ maps sale dates to the year of the transaction. Similarly, I define the middle ring for each T with radius $0.1 < r \leq 0.5$ miles as:

$$R^{mid}_{\ell,T} = \frac{\sum_{i,t} p_{i,t} \cdot \mathbb{1}\{Yr(t) = T\} \times \mathbb{1}\{0.1 < r \leq 0.5\}}{\sum_{i,t} \mathbb{1}\{Yr(t) = T\} \times \mathbb{1}\{0.1 < r \leq 0.5\}} \tag{C.2}$$

Analogously, an outer ring $R^{out}_{\ell,t}$ has radius $0.5 < r \leq 1$ miles. Combining (C.1) and (C.2) gives rise to two sets of DD estimates for each year:

$$Inner\_treatment(T) = (R^{in}_{\ell,T} - R^{out}_{\ell,T}) - (R^{in}_{\ell,-1} - R^{out}_{\ell,-1}) \tag{C.3}$$

$$Middle\_treatment(T) = (R^{mid}_{\ell,T} - R^{out}_{\ell,T}) - (R^{mid}_{\ell,-1} - R^{out}_{\ell,-1}) \tag{C.4}$$

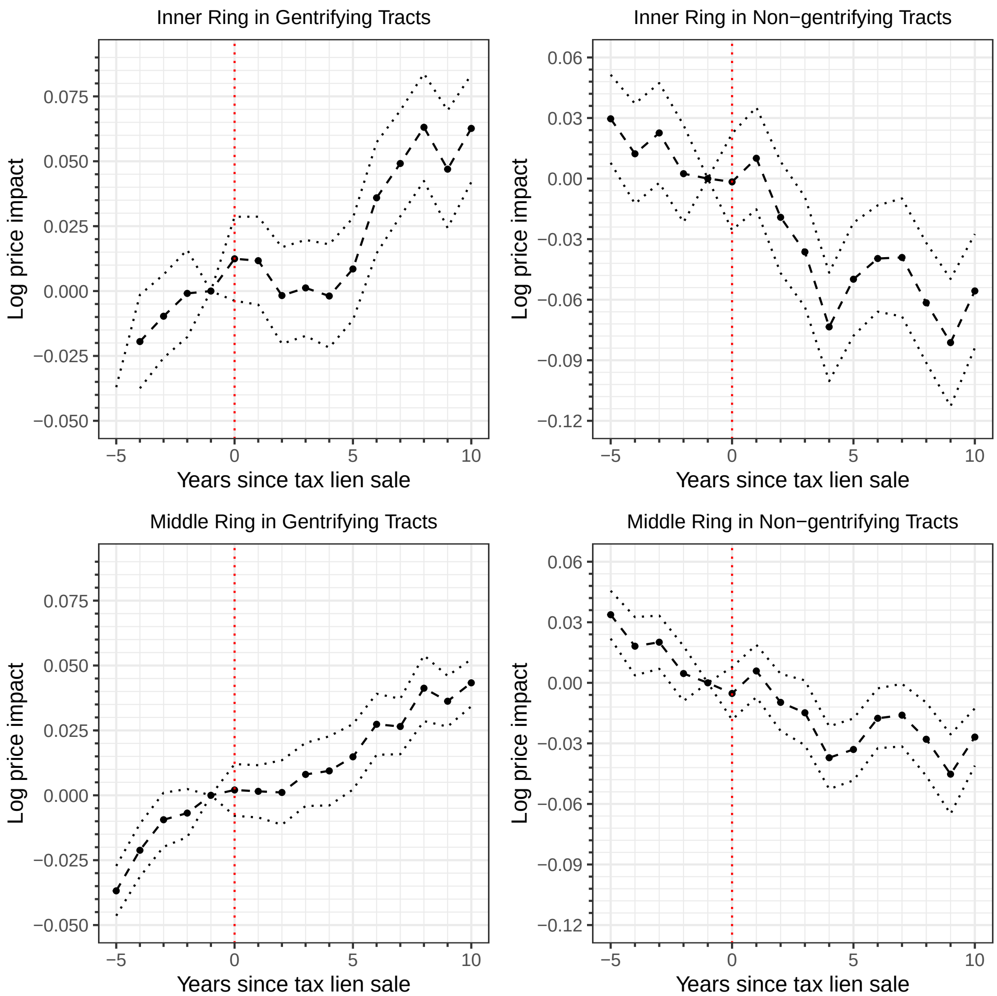
where the differences between rings in T are relative to one year before the tax sale event in $T = 0$.

Using both the inner ring and middle ring treatment definitions, Figure C.1 indicates that prices increase by 5% within six years of a property in a gentrifying area changing hands due to tax sale; the association with neighboring home prices is instead negative in non-gentrifying areas. For the middle ring treatment, Figure C.2 shows similar patterns when I instead split based on initial assessed values for the delinquent properties; pricing effects of tax sales are positive only in above-median home value areas. Additional results in Figure C.3 demonstrate that pooled inner ring estimates from (C.3) averaged over the full sample period decay with distance only within gentrifying tracts, where they remain elevated even at a distance of 0.5 miles.

The pre-existing trends before tax sale events in these figures corroborate the popular narrative that investors target tax-distressed properties located in appreciating housing markets. In the presence of these pre-trends, the results from taking simple differences in ring means are purely descriptive in nature. Accounting for very local time trends and potentially time-varying differences in property characteristics across rings is critical to assess causal effects, which is why I use the foreclosure wave regressions and empirical derivative estimator in Section 4 and Section 5.

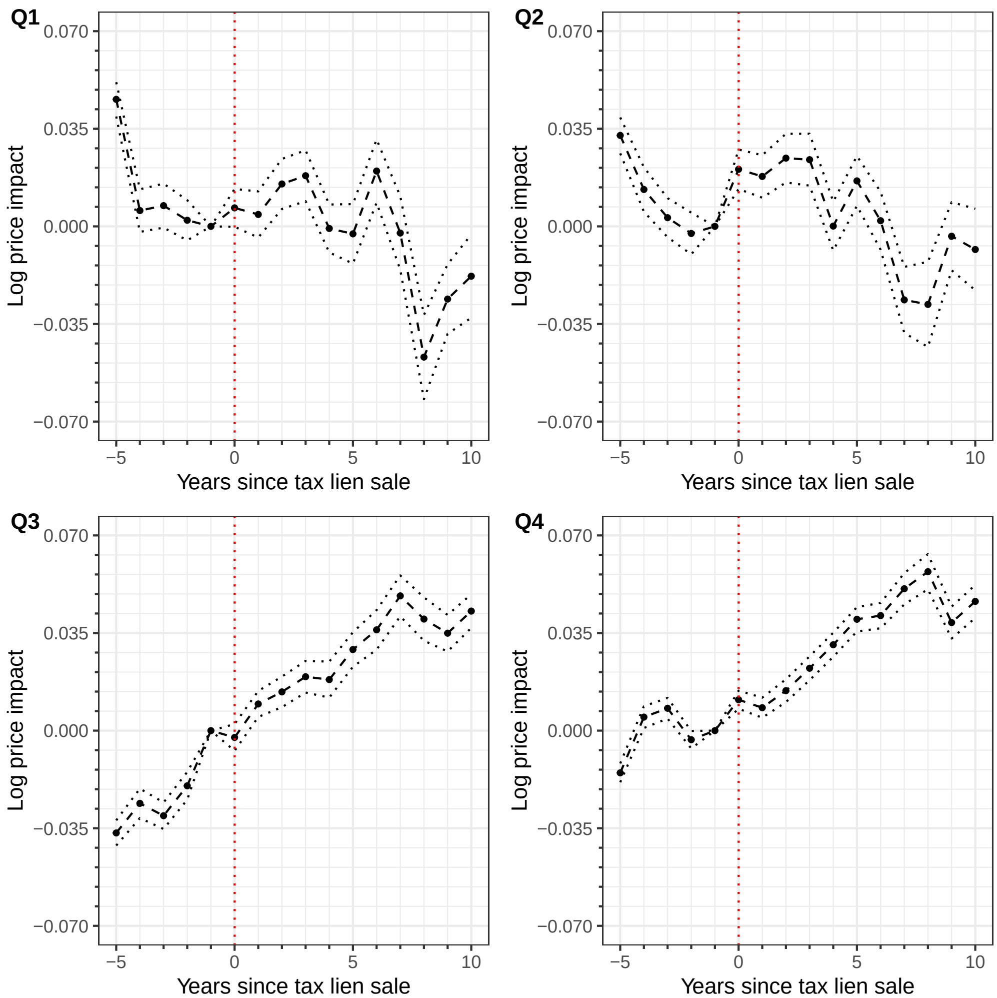
[7] My exposition of this approach follows Appendix B in Diamond & McQuade (2019).

FIGURE C.1. Ring Estimates of Pricing Spillovers by Gentrification Status



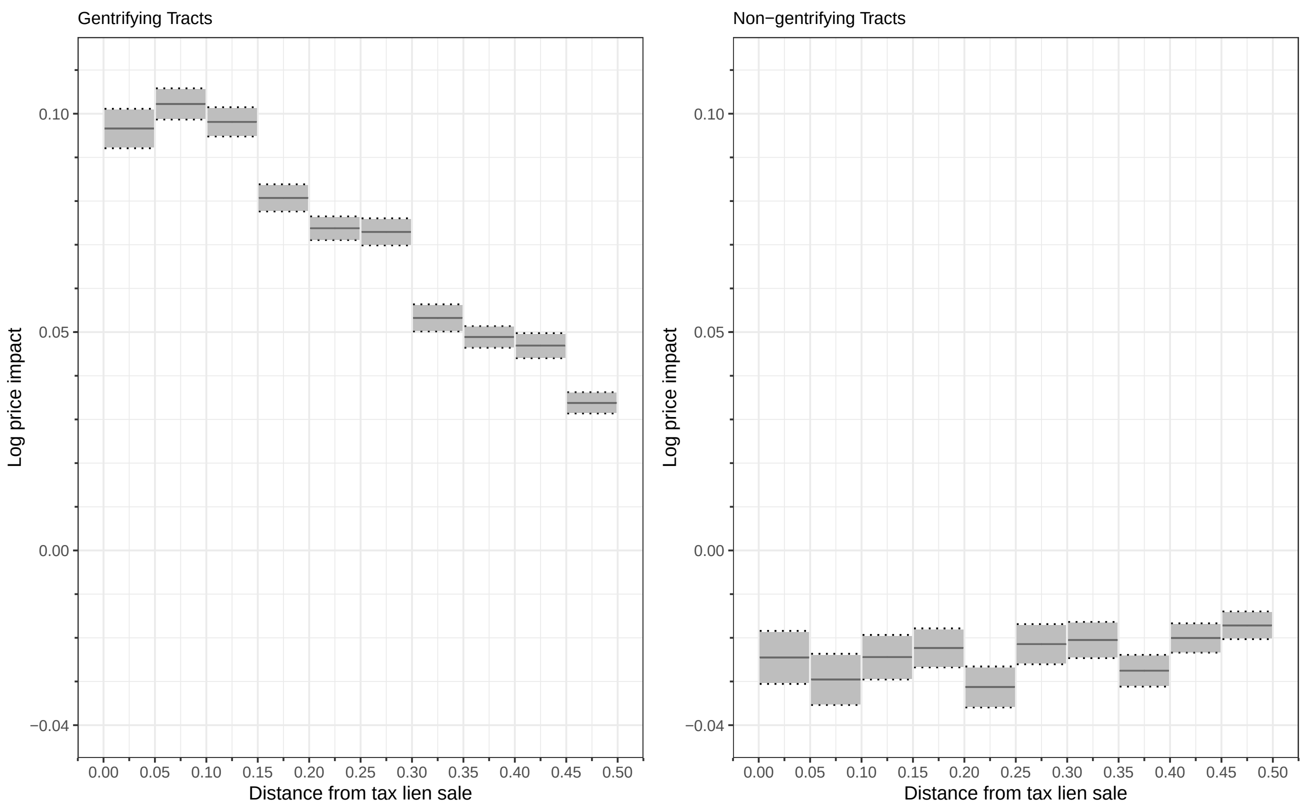
Notes: Each panel shows the evolution of log sale prices per square foot for homes located near a property sold at tax lien auction in Washington, D.C. The tax sale event in these figures is defined as a tax lien sale that results in a title transfer after the six-month redemption period after which a lienholder in D.C. can move to foreclose. While tax-delinquent properties are used to define the ring events, I exclude such properties from the estimation sample. All estimates are relative to one year before a tax lien sale that results in an ownership change, with effects binned at $t = -5$ and $t = +10$ to separate dynamic effects from secular time trends (Schmidheiny & Siegloch 2020). I report separate estimates for gentrifying (left-hand panels) vs. non-gentrifying tracts (right-hand panels) by pooling the tract type definitions applied in Figure 1 and obtained by estimating the population flows model described in Section 3. Inner ring (top panels) refers to estimated effects on prices of homes sold within 0.1 miles of a tax sale property; middle ring (bottom panels) refers to the effect on sale prices 0.1 to 0.5 miles away. Unit prices winsorized at the 1st and 99th percentiles. 95% confidence intervals obtained via 1,000 block bootstrap iterations at the tax sale ring level.

FIGURE C.2. Middle Ring Estimates of Pricing Spillovers by Initial Valuation Tier



Notes: Each panel shows the evolution of log sale prices per square foot for homes located near a property sold at tax lien auction in Washington, D.C. The tax sale event in these figures is defined as a tax lien sale that results in a title transfer after the six-month redemption period after which a lienholder in D.C. can move to foreclose. While tax-delinquent properties are used to define the ring events, I exclude such properties from the estimation sample. All estimates are relative to one year before a tax lien sale that results in an ownership change, with effects binned at $t = -5$ and $t = +10$ to separate dynamic effects from secular time trends (Schmidheiny & Siegloch 2020). I report separate middle ring estimates for neighborhoods organized by the quartile of the tax assessed price per square foot of the distressed property from Zillow ZTRAX as of the start of my sample period. Middle ring refers to the effect on sale prices 0.1 to 0.5 miles away. Unit prices winsorized at the 1st and 99th percentiles. 95% confidence intervals obtained via 1,000 block bootstrap iterations at the tax sale ring level.

FIGURE C.3. Pooled Inner Ring Estimates by Distance to Tax Sale



Notes: Each panel shows the evolution of log home sale prices with respect to distance to a nearby property sold at tax lien auction in Washington, D.C. Following the analysis in the rest of the text, I define a tax sale event as a tax lien sale resulting in a title transfer after the six-month redemption period after which a lienholder in D.C. can move to foreclose. While tax-delinquent properties are used to define the ring events, I exclude such properties from the estimation sample. I report separate estimates for gentrifying (left panel) vs. non-gentrifying tracts (right-hand panel) by pooling the tract type definitions obtained by estimating the population flows model described in Section 3. These are pooled difference-in-differences estimates in that I compare average prices of homes sold in the inner ring within 0.1 miles of a tax sale property relative to an outer ring of properties 0.5 to 1 miles away, and pre vs. post-tax sale event. The estimation procedure follows Butts (2022). Unit prices winsorized at the 1st and 99th percentiles. 95% confidence intervals in gray shaded areas obtained via 1,000 block bootstrap iterations at the tax sale ring level.

C.2 Additional Foreclosure Wave Regression Results

Table C.1. Short-run Spillover Estimates of Tax Sales: Ring Difference-in-Differences

	(1)	(2)	(3)	(4)
$\delta_{C,B} - \delta_{C,A}$	-0.023^{*}	-0.017^{**}	-0.017^{***}	-0.015^{**}
	(0.012)	(0.007)	(0.006)	(0.006)
$\delta_{F,B} - \delta_{F,A}$	0.003	0.003	0.002^{**}	0.002
	(0.003)	(0.002)	(0.001)	(0.002)
$\delta^{p99}_{C,B} - \delta^{p99}_{C,A}$	0.105^{***}	0.059^{***}	0.022^{*}	0.000
	(0.019)	(0.016)	(0.012)	(0.012)
$\delta^{p99}_{F,B} - \delta^{p99}_{F,A}$	0.000	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)
$\delta^{99.5}_{C,B} - \delta^{99.5}_{C,A}$	0.101^{***}	0.058^{***}	0.033^{***}	0.018^{**}
	(0.016)	(0.012)	(0.007)	(0.007)
$\delta^{99.5}_{F,B} - \delta^{99.5}_{F,A}$	0.005^{**}	0.003	0.002	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
$\delta^{99.9}_{C,B} - \delta^{99.9}_{C,A}$	0.084^{***}	0.043^{***}	0.020^{***}	0.009
	(0.014)	(0.010)	(0.007)	(0.008)
$\delta^{99.9}_{F,B} - \delta^{99.9}_{F,A}$	0.002	0.003	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Tract × year FEs	✓	✓	✓	✓
Month FEs	✓	✓	✓	✓
Property controls		✓	✓	✓
Drop old/large			✓	✓
Drop multi-family				✓
N	100,605	88,715	87,472	68,715
Adj. R^2	0.419	0.662	0.762	0.741

Notes: The table reports estimates of tax sale wave regression equation (4.1) using log property price as the outcome, and restricting to arms-length market transactions between individual homebuyers. Each specification allows for piecewise linear effects of $g(\cdot)$ and $h(\cdot)$ over the intervals 0–99th, 99th–99.5th, and 99.5th–99.9th, and $>$ 99.9th percentile. The reported coefficients are estimated differences between the partial derivatives of the "close" function $g(\cdot)$ or the "far" function $h(\cdot)$ before vs. after the occurrence of a tax sale, which renders the estimates comparable to the $t = +1$ estimates in the non-parametric ring analysis conducted in Figure C.1 and Figure C.2. The function $g(\cdot)$ gives a weight to each property within the inner ring equal to 0.1 minus the distance to the tax sale property in miles. $h(\cdot)$ instead adds up with equal weights the number of tax sales within 0.25 miles of the event property. Columns (2), (3), and (4) include controls for the number of bedrooms, bathrooms, floor space, lot size, and a quadratic in property age. Columns (3) and (4) drop properties in the top percentile of size or age, while column (4) drops multi-family properties. Each specification includes a full set of tract × year and month fixed effects. Standard errors clustered at the tract-year level in parentheses. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

C.3 Empirical Derivatives Procedures

As described in Section 4.2, the empirical derivatives (ED) method introduced in Diamond & McQuade (2019), and recently applied by Ganduri & Maturana (2021) to study the spillover effects of property rehabilitation, controls for very local time trends and allows me to trace out how pricing effects continuously evolve with respect to distance from a tax delinquent property. Both papers apply the method to a context where the neighborhoods around the events being analyzed do not overlap. In contrast, events like tax sales and mortgage foreclosures are highly geographically clustered, which means some adjustments must be made in the tuning parameters to account for differences in event sparsity. Here I summarize how I apply the method to the tax sale context. Interested readers should refer to Section IV and the Appendix of Diamond & McQuade (2019) for more details on the algorithm.

The ED method supposes house prices $p_{i,t}$ in a neighborhood of tax sale property S follow:

$$\log(p_{i,t}) = m_S(d_i, \tau_i) + \phi_S(d_i, \theta_i) + \gamma_S(\theta_i, t_i) + \varepsilon_{i,t} \tag{C.5}$$

where property i relates to S in the polar coordinates plane (d, θ), with d the distance and θ the direction to the nearest tax sale property. The non-parametric function of interest is $m_S(\cdot)$ which captures the spillover effect of S on i located d_i miles away and sold τ_i years after the tax sale. The functions $\phi_S(\cdot)$ and $\gamma_S(\cdot)$ control for very local neighborhood effects along both directions in the polar plane and direction-time trends, respectively.

The estimation of $m_S(\cdot)$ proceeds in two steps. First, the algorithm estimates an empirical partial derivative of the house price with respect to distance d for each transaction, defined by the triple (d_i, θ_i, t_i). Second, a kernel regression smooths the partial derivative estimates to obtain a smooth surface (like those presented in Figure 8 and Figure 9) of how house prices vary with respect to d and τ; this is equivalent to the partial derivatives of the function $m_S(\cdot)$.

The first step involves searching within a bowtie-shaped area within a larger ring of radius r for pairs of houses which are slightly closer or farther from the tax sale property by some finite distance tolerance $\delta > 0$. Then for each price pair, the partial derivative of $p_{i,t}$ with respect to d and holding θ_i and t_i fixed is:

$$\frac{\log(p_{d-\delta,\theta,t}) - \log p_{d+\delta,\theta,t}}{2\delta}, \quad \text{for } \delta > 0 \tag{C.6}$$

The algorithm then stacks up the differences computed via (C.6) for properties close to the tax sale property S and computes them relative to one year before the tax sale event. Comparing outcomes for two "close" house pairs differences out the time trend $\gamma_S(\cdot)$, and taking estimates relative to $t_i = -1$ takes out the time-invariant micro-geography effects captured by $\phi_S(\cdot)$.

I define the tax sale event in my baseline estimation as a tax lien sale resulting in a title transfer after the statutory redemption period ends (i.e. > 6 months after the lien auction in Washington, D.C.). In practice, depending on the neighborhood, there could be few or many properties around the tax sale property which transact at (θ_i, t_i) but with different distances. Hence, the estimates are influenced by the choice of tuning parameters governing the size of the bowtie and how the partial derivatives are smoothed as one moves within the ring along the θ and t directions.

To summarize, the researcher chooses the following six parameters to implement this technique:

- $h_{r,n}$: **kernel smoothing in distance (miles).**

- $h_{t,n}$: **kernel smoothing in time (years).**
- g_n^t: **bowtie search area width in time.**
- g_n^θ: **bowtie search area width in polar distance.**
- κ_n: **maximum number of house pairs included in the bowtie.**
- r: **ring radius within which to trace out the derivative.**

In Figure C.4 and Figure C.5, I check how the estimated 3-D price surfaces from Figure 8 vary according to parameter choices using the six sets of parameters listed in Table C.2. The list in column (I) of Table C.2 corresponds to the set of parameters used by Diamond & McQuade (2019); the list in column (VI) is used by Ganduri & Maturana (2021). Column (III) reflects the parameter set I use in my baseline estimation of Figure 8 and Figure 9. For neighboring house prices as the outcome, I find the derivative estimates are relatively invariant to the bowtie dimensions g_n^t and g_n^θ, or the maximum number of housing price pairs, so I fix those at the values used in the previous two papers employing this method.

However, due to significant racial segregation at very fine neighborhood levels, the house pair parameter κ_n is important for measuring the effects on buyer or seller race as an outcome. This is because without enough house pairs there will not be any demographic variation between pairs $\pm\delta$ away from a tax sale property (e.g. all properties will be owned by either white or non-white taxpayers), leading to corner solutions non-convergence. Simply increasing κ_n will not resolve the problem when thin local markets with few price pairs are also among the most segregated, as is the case in a city like D.C. I address this issue in my analysis of demographic spillovers in Section 5 by using a semi-parametric event study design which discretizes the empirical derivatives approach.

While the general shape of the 3-D price surface is the same across parameter sets in Figure C.4 and Figure C.5, there are some notable differences when I impose a larger ring radius r (Column I) or substantially lower the time smoothing parameter $h_{t,n}$ (Column VI). There are intuitive economic reasons for this that guide my baseline tuning parameter choices. Increasing the ring radius and smoothing over a larger distance dimension as in Column I results in a flatter pricing surface for gentrifying areas, since large differences between pricing pairs are averaged together with small differences between pricing pairs. As Figure C.3 illustrates, average price differences between far and close properties to a tax sale event rapidly decay with respect to distance. Diamond & McQuade (2019) use larger search distance parameters because they study the local effects of affordable housing constructed under the Low Income Housing Tax Credit (LIHTC) program that by design selects project sites not proximal to existing affordable housing developments.

In Ganduri & Maturana (2021) rehabs are not geographically clustered, and the redevelopment process is fast (median 70 days); consequently those authors use large distance smoothing and small time smoothing parameters. Applying their parameters in Column VI leads to non-monotonic estimates of spillovers of tax sales, as housing markets around tax sales are generally thick, and even after the redemption period ends, redevelopment can be a long process due to legal proceedings involved in transferring the tax deed and receiving permits needed to convert from single-family to multi-family or commercial units (see Section 2.1 for legal background).

C.4 Additional Results Using Empirical Derivatives

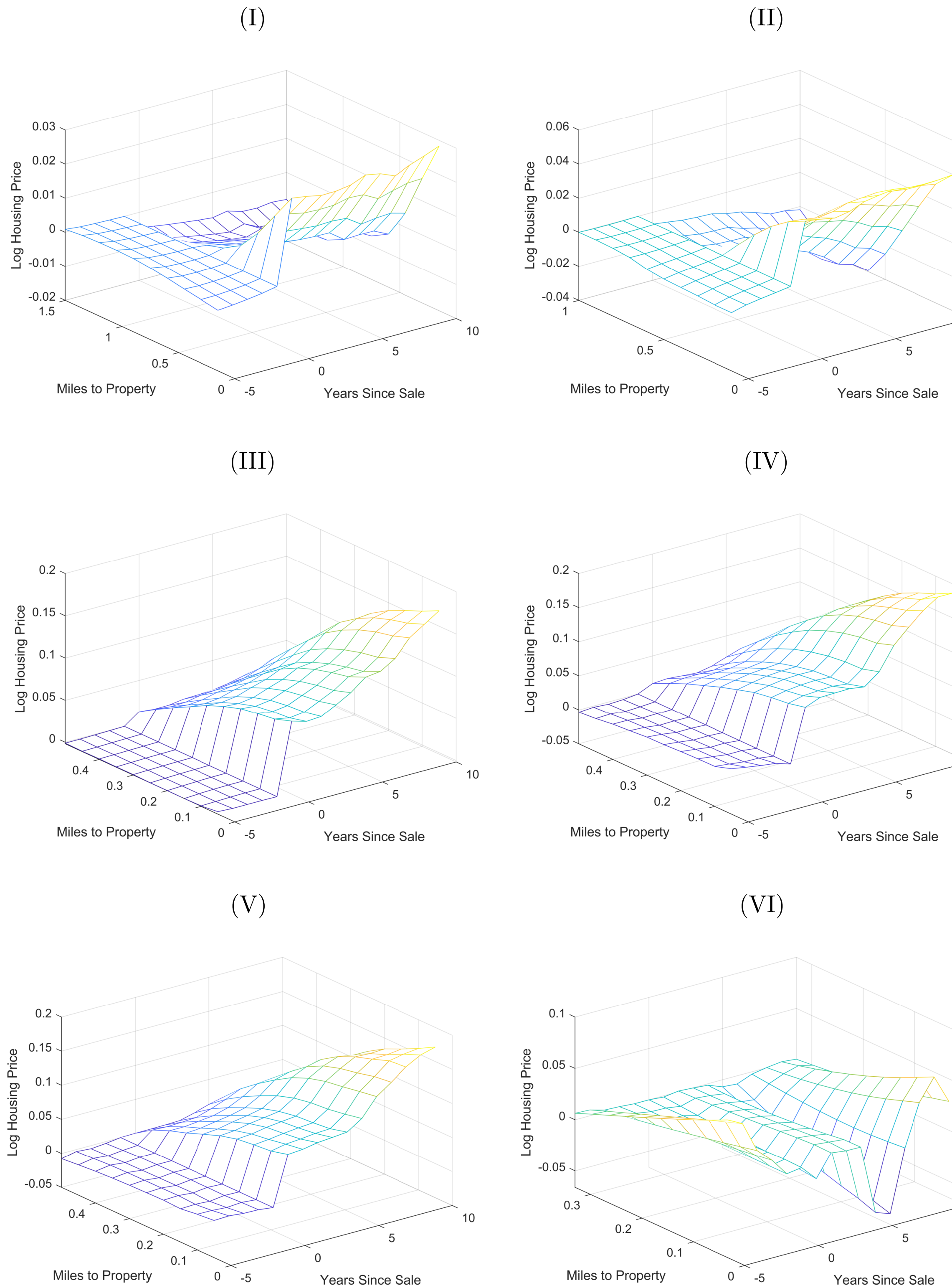
Here I summarize results on other potential sources of heterogeneity in the local pricing effects of tax sales using the empirical derivatives method. To summarize:

Table C.2. Alternative Tuning Parameters for Empirical Derivatives Estimator

	(I)	(II)	(III)	(IV)	(V)	(VI)
Smoothing parameters						
$h_{r,n}$ (smoothing in miles)	0.300 mi.	0.200 mi.	0.125 mi.	0.125 mi.	0.100 mi.	0.250 mi.
$h_{t,n}$ (smoothing in years)	5 years	5 years	5 years	3 years	3 years	1.5 years
Bowtie dimensions						
g_n^t (width in years)	1.6 years	1.6 years	1.6 years	1.6 years	1.6 years	1.6 years
g_n^θ (width in polar distance)	0.4	0.4	0.4	0.4	0.4	0.4
Sample selection						
κ_n (# price pairs)	5 pairs	5 pairs	5 pairs	8 pairs	5 pairs	5 pairs
ℓ_n (excluded zone)	0.01 mi.	0.01 mi	0.01 mi.	0.01 mi.	0.01 mi.	0.01 mi.
r (ring radius)	1.5 mi.	1 mi.	0.5 mi.	0.5 mi.	0.5 mi.	0.33 mi.

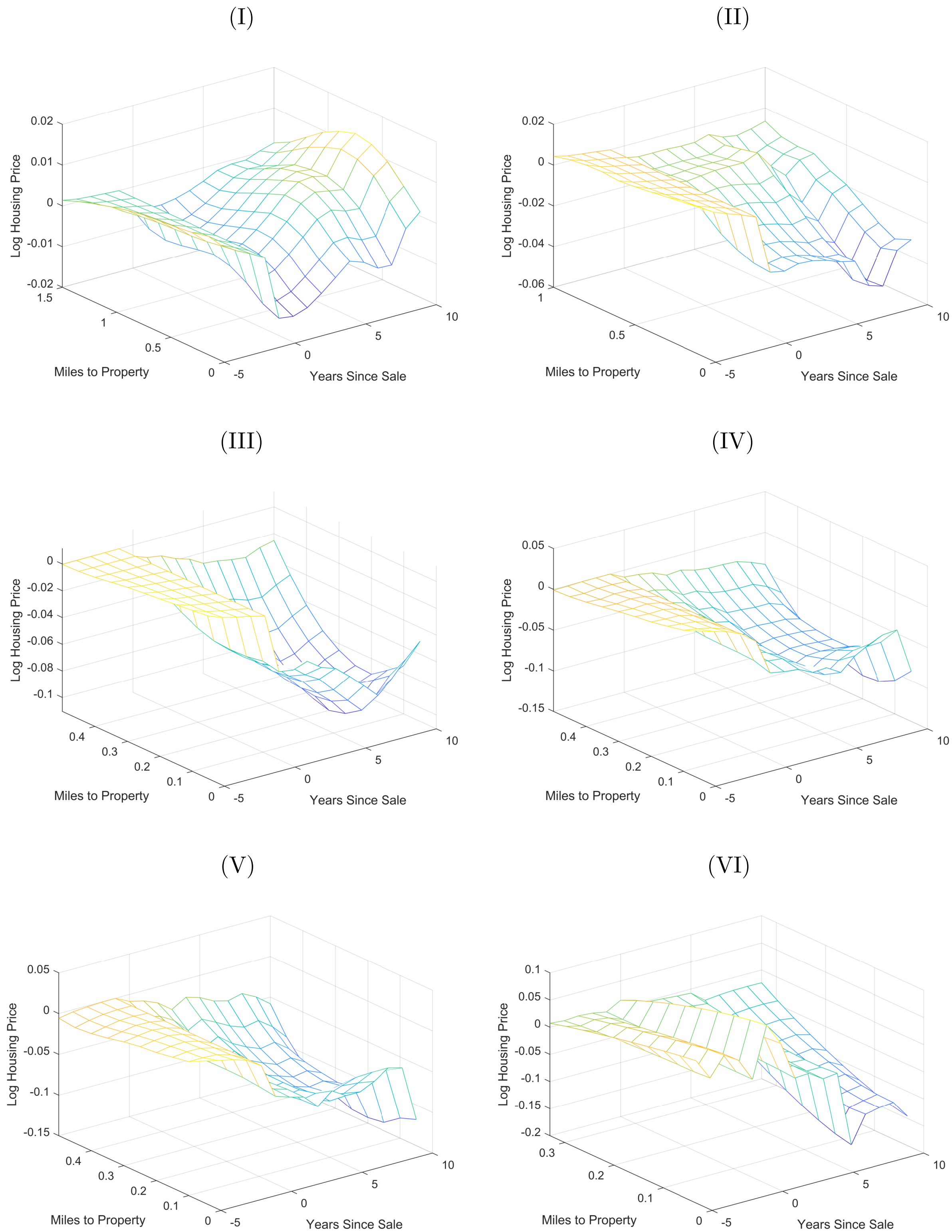
Notes: Parameter notation follows the presentation of the empirical derivative method in Section IV.B of Diamond & McQuade (2019), and column (I) corresponds to the set of parameters used in that paper. Column (VI) corresponds to the parameters used in Ganduri & Maturana (2021). I use the set of parameters in column (III) in establishing my main results. A distinguishing feature of my setting relative to others which have used this method is that tax sales are more numerous and frequently occur in tiny geographic pockets, which rationalizes my use of a smaller distance smoothing parameter and ring radius.

FIGURE C.4. Empirical Derivatives for Different Tuning Parameters (Gentrifying)



Notes: Each panel shows the continuous evolution of log sale prices for homes located near a property sold at tax lien auction in Washington, D.C. Estimates obtained by applying the empirical derivatives estimator of Diamond & McQuade (2019), where the panel numbers map to the set of tuning parameters in the corresponding column of Table C.1. Tax lien sale properties are excluded from the estimation sample to avoid reverse causality. All estimates are relative to the year before the event of a tax lien sale to a buyer in a gentrifying tract, with gentrification defined using the two-stage model in Section 3.

FIGURE C.5. Empirical Derivatives for Different Tuning Parameters (Non-gentrifying)



Notes: Each panel shows the continuous evolution of log sale prices for homes located near a property sold at tax lien auction in Washington, D.C. Estimates obtained by applying the empirical derivatives estimator of Diamond & McQuade (2019), where the panel numbers map to the set of tuning parameters in the corresponding column of Table C.1. Tax lien sale properties are excluded from the estimation sample to avoid reverse causality. All estimates are relative to the year before the event of a tax lien sale to a buyer in a non-gentrifying tract, with gentrification defined using the two-stage model in Section 3.

- Figure C.4 demonstrates that purchases of tax claims by non-institutional investors result in different pricing trends than the institutional investor purchases highlighted in see Figure 8 and Table 4. While there are strong positive price spillovers in (Panel C) for opportunistic investors in recently gentrifying neighborhoods, there are no clear effects based on whether the neighborhood was already gentrifying. Moreover, given the small number of sales to non-institutional investors (roughly one-third of the overall tax lien sample for D.C.) the standard errors on these point estimates are much wider.

- Table C.2 reports point estimates and standard errors from the ED method applied to all tax sale neighborhoods pooled together. This results in a statistically flat pricing surface, albeit with prices trending upward post-tax sale, indicating that the heterogeneous effects between gentrifying and non-gentrifying areas washes out in the aggregate sample.

- Table C.3 reports ED point estimates and standard errors, as in Table 4, but for tax sale in recently non-gentrifying tracts. There is a clear negative pre-trend at close distances ($<$ 0.3 miles) to a tax-delinquent property, but a sharp negative drop in prices within a year after the redemption period on a tax lien ends. The inability of the ED method to remove local pre-trends for the case of distressed neighborhoods may reflect the influence of strong cross-contamination treatment effects, whereby overlapping non-gentrifying tax sale rings magnify the impact of each additional tax foreclosure event. This contamination mechanism is reflected in the foreclosure wave regression analysis of Section 4.2, where the average effect of a tax sale even at relatively "far" distances ($>$ 0.25 miles away) is negative.

D Pricing Foreclosure Options: The Repeat Distress Index

In this appendix I document how the value of claims to tax-distressed properties varies over time using a method which I call the "repeat distress index." Like the repeat sales methodology which is commonly used to generate a series of quality-adjusted price levels and compare variation across local real estate markets, I use multiple transactions of the same property to take out the property fixed effect. However, unlike repeat sales, in this context the transaction is a premium bid on the property at a tax lien auction, so the repeat event is a tax delinquency. Hence, the selection into the sample of repeat transactions is *negative* rather than positive.

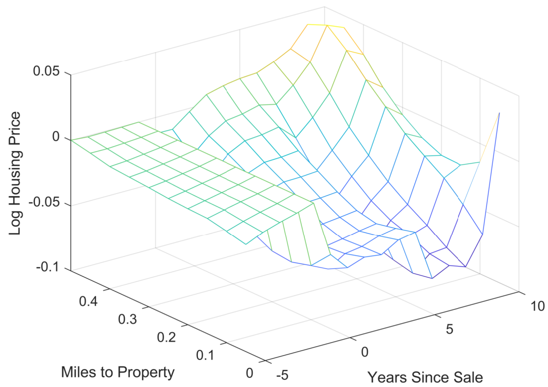
As discussed in the legal background of Section 2 and Appendix A, in 39 out of the 50 states + D.C. local governments auction off claims to tax-delinquent properties via the premium bid method, where the starting bid is the total tax debt (overdue payments + interest + penalties).[8] In the event of foreclosure, the lienholder can "credit bid" by bidding the lien amount plus any interest or penalties accrued on the lien during the redemption period. Thus this initial premium bid is a reasonable proxy for the market price of a foreclosure option, regardless of whether the delinquency results in an eventual foreclosure.

My tax sale data contain square-suffix-lot (SSL) combinations for each attached lien, which offers one way to track liens against the same property over time. However, since properties can experience multiple delinquency events across different owners and uses, the SSL does not allow me to directly track identical properties with multiple liens over time. This not necessarily an issue for

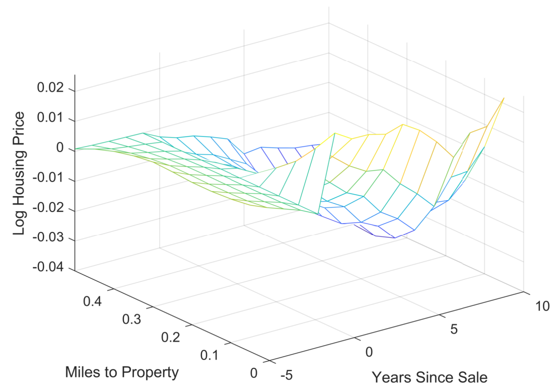
[8]There are a few exceptions to the rule that the starting bid is set to the break even point from the tax authority's perspective. For instance, Ohio the starting bid is two-thirds of assessed value; it is \$800 in Pennsylvania, and 100% of appraised value in Wisconsin. All three of these states are deed states.

FIGURE C.6. Estimates of Tax Sale Pricing Spillovers (Non-institutional Lien Buyers)

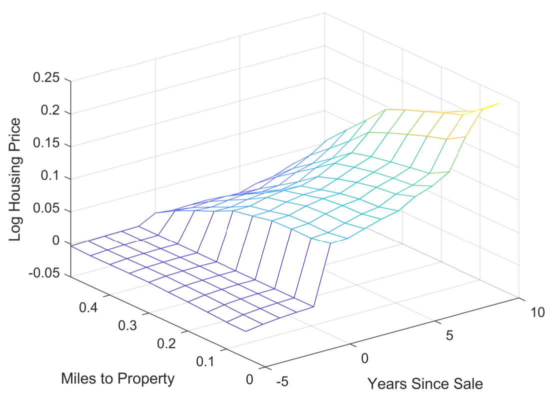
A. Sales in Previously Gentrifying Tracts



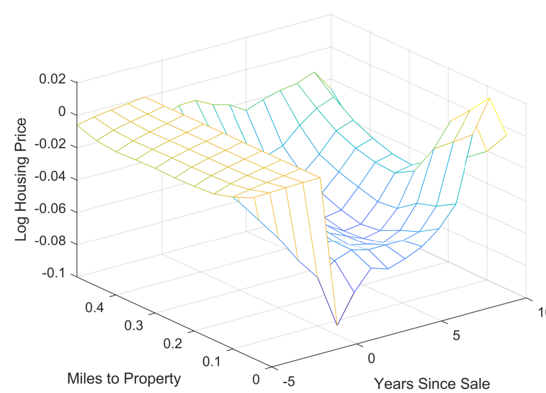
B. Sales in Previously Non-gentrifying Tracts



C. Sales in Recently Gentrifying Tracts



D. Sales in Recently Non-gentrifying Tracts



Notes: Each panel shows the continuous evolution of log sale prices for homes located within a 0.5 mile radius of a property sold at tax lien auction in Washington, D.C. Estimates obtained by applying the empirical derivatives estimator of Diamond & McQuade (2019) and excluding all properties within 0.01 miles of the tax sale property. Tax lien sale properties are excluded from the estimation sample to avoid reverse causality. All estimates are relative to the year before the event of a tax lien sale to an individual buyer. I identify institutional buyers using the keywords: “LLC”, “FUND”, “INC”, “BANK”, “REALTY”, “PARTNERS”, “CAPITAL”, “TRUST”, “CORPORATION”, “PLLC.” Individuals exclude institutional buyers and non-profit buyers identified using the keywords: “PRAYER”, “CHURCH”, “COMMUNITY”, “FAITH”, “UNIVERSITY”, “COLLEGE”, “SCHOOL”, “BAPTIST”, “FOUNDATION”, “GOVERNMENT”, “EMBASSY”, “CENTER”, “COOPERATIVE”, “FRIENDSHIP”, “MINISTRIES”, “FEDERAL”, “REHABILITATION.” Panel A reports estimates for gentrifying tracts, while Panel B does the same for non-gentrifying tracts according to the Census tract type definitions obtained by estimating the population flows model described in Section 3 over the period 1990-2005. Panels C and D conduct the same exercise as in Panels A and B, respectively, except with gentrifying areas identified by running the classification model over the later period in 2005-2019.

Table C.2. Pricing Effects of Tax Sales on Nearby Properties in All Tracts

Distance (mi.) →	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Event time (years) ↓	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
−5	0.0067*	0.0047	0.0038	0.0031	0.0026	0.0021	0.0010	0.0001	−0.0006	−0.0006
	(0.0040)	(0.0033)	(0.0030)	(0.0028)	(0.0026)	(0.0024)	(0.0022)	(0.0019)	(0.0015)	(0.0010)
−4	0.0041*	0.0030	0.0025	0.0021	0.0017	0.0013	0.0006	0.0000	−0.0005	-0.0006
	(0.0024)	(0.0020)	(0.0019)	(0.0017)	(0.0016)	(0.0015)	(0.0013)	(0.0012)	(0.0009)	(0.0006)
−3	0.0027*	0.0020	0.0018	0.0015	0.0012	0.0009	0.0005	0.0000	−0.0003	−0.0004
	(0.0015)	(0.0013)	(0.0012)	(0.0011)	(0.0011)	(0.0010)	(0.0009)	(0.0008)	(0.0006)	(0.0004)
−2	0.0016*	0.0012	0.0010	0.0009	0.0007	0.0006	0.0003	0.0000	−0.0002	−0.0002
	(0.0008)	(0.0007)	(0.0007)	(0.0006)	(0.0006)	(0.0005)	(0.0005)	(0.0004)	(0.0003)	(0.0002)
1	0.0138	0.0100	0.0073	0.0054	0.0028	0.0012	−0.0005	−0.0011	−0.0015	−0.0009
	(0.0118)	(0.0104)	(0.0097)	(0.0092)	(0.0088)	(0.0084)	(0.0078)	(0.0069)	(0.0056)	(0.0037)
2	0.0160	0.0089	0.0055	0.0034	0.0009	−0.0006	−0.0025	−0.0030	−0.0030	−0.0017
	(0.0120)	(0.0106)	(0.0098)	(0.0093)	(0.0089)	(0.0086)	(0.0080)	(0.0072)	(0.0059)	(0.0040)
3	0.0163	0.0083	0.0037	0.0011	−0.0014	−0.0031	−0.0052	−0.0057	−0.0054	−0.0032
	(0.0118)	(0.0104)	(0.0097)	(0.0092)	(0.0088)	(0.0084)	(0.0078)	(0.0069)	(0.0056)	(0.0037)
4	0.0191	0.0089	0.0035	0.0002	−0.0025	−0.0046	−0.0072	−0.0080	−0.0076	−0.0046
	(0.0120)	(0.0106)	(0.0098)	(0.0093)	(0.0089)	(0.0086)	(0.0080)	(0.0072)	(0.0059)	(0.0040)
5	0.0210	0.0088	0.0027	−0.0010	−0.0039	−0.0066	−0.0099	−0.0113	−0.0110	−0.0073
	(0.0122)	(0.0108)	(0.0101)	(0.0096)	(0.0092)	(0.0089)	(0.0084)	(0.0076	(0.0064)	(0.0044)
10	0.0555***	0.0388*	0.0301	0.0211	0.0156	0.0110	0.0051	0.0017	0.0000	−0.0009
	(0.0209)	(0.0199)	(0.0194)	(0.0179)	(0.0174)	(0.0171)	(0.0166)	(0.0158	(0.0141)	(0.0101)

Notes: This table presents average log house prices at various distances in event time relative to the one year before a tax lien sale in all tracts. Note that estimates at event time = 0 correspond to cases where the lien is potentially held within the six-month redemption period where foreclosure by the lienholder is not yet possible, and consequently I suppress these from the table due to a lack of clear interpretation. Estimates computed using the empirical derivative method of Diamond & McQuade (2019) described in the main text and Appendix C.3. Tuning parameters follow those in column (III) of Table C.1. Standard errors computed using the block bootstrap method with 500 sample draws, where sampling is carried over neighborhoods corresponding to tax lien properties. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table C.3. Pricing Effects of Tax Sales on Nearby Properties in Recently Non-gentrifying Tracts

Distance (mi.) → Event time (years) ↓	0.05 (1)	0.1 (2)	0.15 (3)	0.2 (4)	0.25 (5)	0.3 (6)	0.35 (7)	0.4 (8)	0.45 (9)	0.5 (10)
−5	0.0116**	0.0101**	0.0088**	0.0074*	0.0063*	0.0050	0.0031	0.0012	0.0002	−0.0002
	(0.0051)	(0.0047)	(0.0043)	(0.0040)	(0.0038)	(0.0035)	(0.0032)	(0.0028)	(0.0023)	(0.0014)
−4	0.0075**	0.0063**	0.0056**	0.0048*	0.0041*	0.0032	0.0020	0.0007	0.0000	−0.0002
	(0.0032)	(0.0029)	(0.0027)	(0.0025)	(0.0024)	(0.0022)	(0.0020)	(0.0017)	(0.0014)	(0.0009)
−3	0.0050**	0.0041**	0.0037**	0.0032*	0.0027*	0.0021	0.0013	0.0005	0.0000	−0.0001
	(0.0021)	(0.0019)	(0.0018)	(0.0017)	(0.0016)	(0.0014)	(0.0013)	(0.0011	(0.0009)	(0.0006)
−2	0.0027**	0.0022**	0.0020**	0.0018*	0.0015*	0.0012	0.0007	0.0003	0.0000	−0.0001
	(0.0012)	(0.0011)	(0.0010)	(0.0009)	(0.0009)	(0.0008)	(0.0007)	(0.0006)	(0.0005)	(0.0003)
1	−0.0461**	−0.0486***	−0.0523***	−0.0533***	−0.0528***	−0.0487***	−0.0423***	−0.0342***	−0.0248***	−0.0123**
	(0.0190)	(0.0169)	(0.0155)	(0.0146)	(0.0138)	(0.0132)	(0.0124)	(0.0112	(0.0093)	(0.0063)
2	−0.0424**	−0.0524***	−0.0577***	−0.0591***	−0.0584***	−0.0541***	−0.0476***	−0.0388***	−0.0280***	−0.0138**
	(0.0195)	(0.0173)	(0.0161)	(0.0152)	(0.0144)	(0.0138)	(0.0131)	(0.0120	(0.0101)	(0.0069)
3	−0.0499**	−0.0609***	−0.0676***	−0.0694***	−0.0684***	−0.0639***	−0.0572***	−0.0475***	−0.0354***	−0.0188**
	(0.0190)	(0.0169)	(0.0155)	(0.0146)	(0.0138)	(0.0132)	(0.0124)	(0.0112	(0.0093)	(0.0063)
4	−0.0535**	−0.0672***	−0.0747***	−0.0769***	−0.0759***	−0.0713***	−0.0645***	−0.0542***	−0.0409***	−0.0221**
	(0.0195)	(0.0173)	(0.0161)	(0.0152)	(0.0144)	(0.0138)	(0.0131)	(0.0120	(0.0101)	(0.0069)
5	−0.0624**	−0.0780***	−0.0858***	−0.0877***	−0.0865***	−0.0821***	−0.0753***	−0.0644***	−0.0497***	−0.0280**
	(0.0204)	(0.0184)	(0.0172)	(0.0162)	(0.0155)	(0.0149)	(0.0142)	(0.0132	(0.0113)	(0.0079)
10	−0.0503	−0.0740**	−0.0832**	−0.0918***	−0.0908***	−0.0849***	−0.0760**	−0.0625**	−0.0453*	−0.0252
	(0.0369)	(0.0356)	(0.0347)	(0.0321)	(0.0312)	(0.0305)	(0.0299)	(0.0288)	(0.0259)	(0.0188)

Notes: This table presents average log house prices at various distances in event time relative to the one year before a tax lien sale in non-gentrifying tracts. Note that estimates at event time = 0 correspond to cases where the lien is potentially held within the six-month redemption period where foreclosure by the lienholder is not yet possible, and consequently I suppress these from the table due to a lack of clear interpretation. Estimates computed using the empirical derivative method of Diamond & McQuade (2019) described in the main text and Appendix C.3. Tuning parameters follow those in column (III) of Table C.1. Standard errors computed using the block bootstrap method with 500 sample draws, where sampling is carried over neighborhoods corresponding to tax lien properties. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

hedonic indexing methods, which use a set of possibly time-varying observables to price properties in the cross-section. An hedonic approach would, however, require strong assumptions about the underlying functional form for foreclosure option values.

At the same time, repeat sales indices have well-known selection bias issues; properties undergoing multiple delinquency events are likely fundamentally different from properties where the owner remains current on their tax bill. To resolve these tradeoffs, I adopt a hybrid hedonic-approach in the spirit of McMillen (2012) and Fang et al. (2015) that transforms the time fixed effects in the following regression to estimate a lien purchase price index:

$$\log B_{i,t} = \delta_t + \gamma_m + \widetilde{\alpha}_i + \beta' \cdot \mathbf{X_{i,t}} + \epsilon_{i,t} \tag{D.1}$$

$$B_t = \exp(\delta_t) \tag{D.2}$$

where i indexes a property with an attached lien certificate sold in year t, δ_t are year fixed effects, and γ_m are dummies for the month of the tax sale. The property type fixed effects $\widetilde{\alpha}_i$ control for all time-invariant observed or unobserved characteristics of the transacted property type. $B_{i,t}$ is the final bid the lienholder pays.

I identify the $\widetilde{\alpha}_i$ by matching transactions on geolocation information and other features to determine "uniqueness" of a transaction. I consider four variations of this method, with uniqueness defined with decreasing stringency as one goes down the following list:

1. **Address $\times$ square $\times$ lot fixed effects:** I assign two liens to the same panel id if they are attached to the same standardized address of the concatenated format: number + street name, and share the same lot and square in the SSL. This method results in an almost identical set of repeat liens as the below Method 2 (41.4% of liens included).

2. **Address $\times$ lot fixed effects:** Two liens share a panel id if they are attached to the same standardized address of the concatenated format: number + street name and share the same lot in the SSL. This panel id is more stringent than Method 3 below in that it accounts for cases where the parcel has been divided into separate tax units (e.g. different floors in a townhouse), even if the building and street address is the same (41.7% of liens included).

3. **Address-level fixed effects:** Two liens share a panel id if they have the same standardized address of the concatenated format: number + street name. This method effectively treats two units within a building as the same (56.7% of liens included).

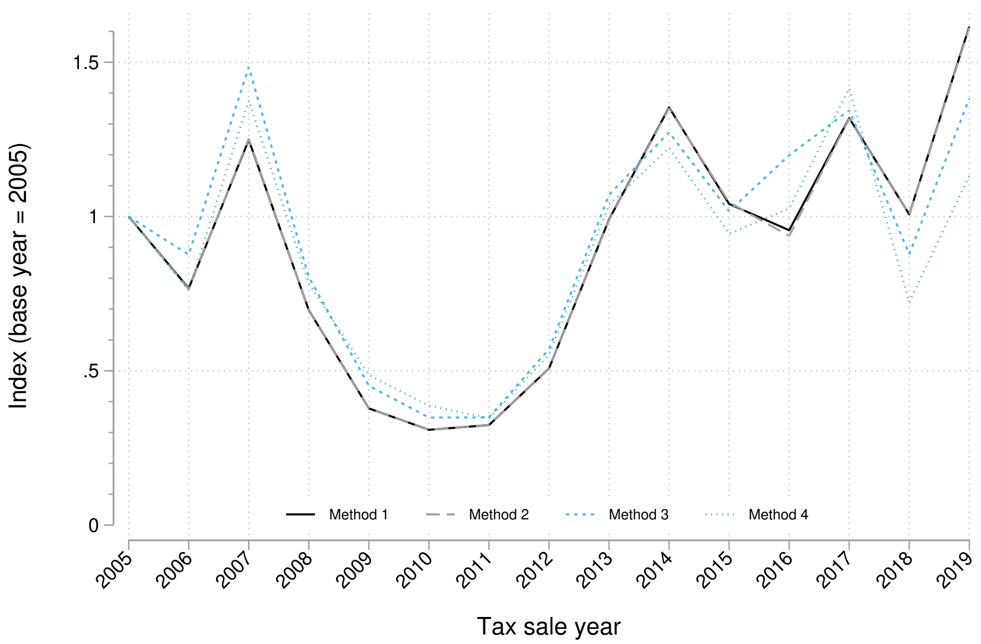
4. **Block-level fixed effects:** I assign two liens to the same panel id if they share the same CoreLogic Tax block number. This is roughly equivalent to including city block fixed effects, or controlling for time-invariant local neighborhood characteristics (95.8% of liens included).

In the regression, the vector $\mathbf{X_{i,t}}$ includes a polynomial in land area and floor space, the number of floors in the building, and the unit floor for apartments. To the extent that each method could assign two distinct but adjacent properties to the same panel id, controlling for $\mathbf{X_{i,t}}$ accounts for further differences due to the height and size which may be relevant to investors' valuation.

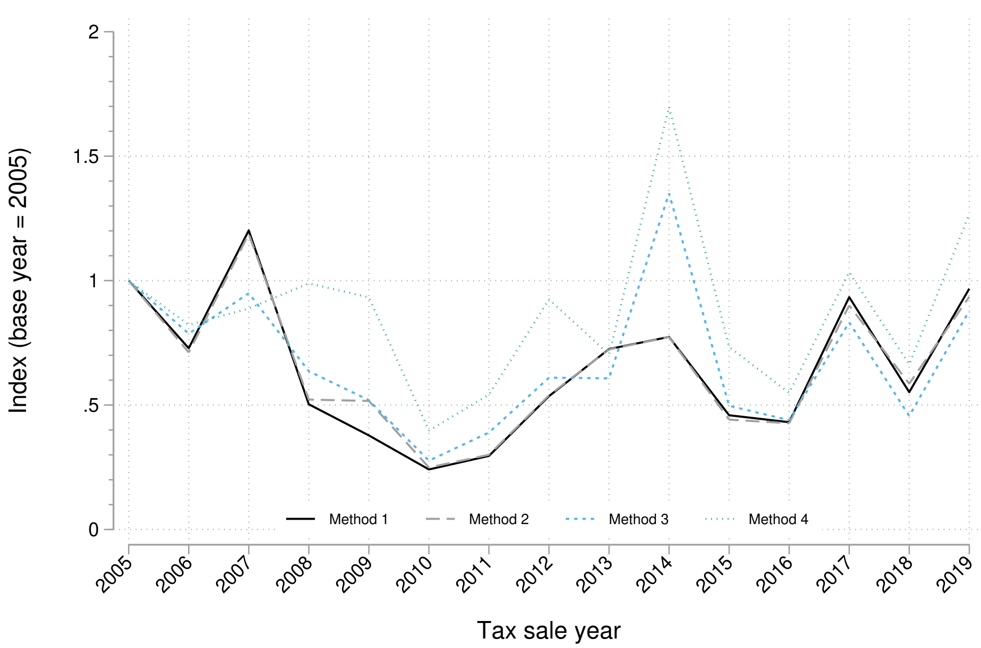
In estimating these four versions of the model in (D.1) and (D.2), I drop lien sales which (i) occur outside known auction dates, and (ii) for which the surplus bid equals zero. Such liens are over-the-counter sales involving relatively undesirable properties for which the tax authority was unable to sell ownership claims at auction. In this way, I estimate the price indices on readily redeployable properties.

FIGURE D.1. Repeat Distress Index for Tax Liens: Total Bid Value

A. Institutional Lien Buyers

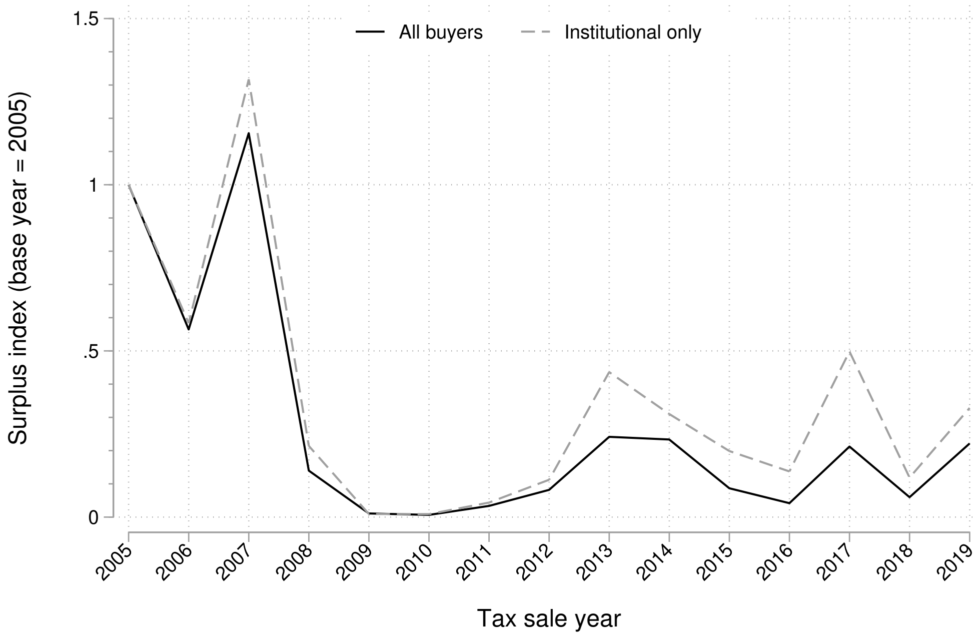


B. Individual Lien Buyers



Notes: This figure plots my repeat distress indices using the matching estimator method implemented via equations (D.1) and (D.2). Each method "matches" the same property across different tax sale events using a different set of conditioning variables related to the property location to identify panel fixed effects. Method 1 is the most stringent definition of a unique property, while Method 4 is the least stringent. Full details underlying the matching in each method in the text to Appendix D. Panel A restricts to tax liens sold to institutional investors, while Panel B restricts to those sold to individual investors. I use the same keywords in the investor's name as in the rest of the paper to separate for-profit institutions from individuals. All bid values in real 2012 dollars, converted from nominal terms using the PCE deflator.

FIGURE D.2. Repeat Distress Index for Tax Liens: Surplus Value



Notes: This figure plots my repeat distress indices for surplus bids using the matching estimator Method 3 implemented via equations (D.3) and (D.4). Method 3 matches based on address-level fixed effects. I use the same keywords in the investor's name as in the rest of the paper to separate for-profit institutions from individuals. All bid values in real 2012 dollars, converted from nominal terms using the PCE deflator.

Figure D.1 separately plots the resulting repeat distress indices for each method and for institutional lien buyers (Panel A) and individual lien buyers (Panel B). Each index series uses 2005 as the base year. All four methods closely track each other – above 90% correlation between each pairwise comparison of series. The lien valuations reflect the general housing cycle; values peak during the foreclosure crisis 2006-2007, and collapse during the recession years before sharply recovering in the institutional investor segment of the market. The differences between methods are informative about how selection of delinquent properties varies over the housing cycle. Generally, the price levels from Method 1 are lower than the other methods, especially in the first half of the sample when the market is thicker. This is the opposite direction of the selection bias effect one would expect from implementing a standard repeat sales/matching estimator method. The reason is that as we move towards a more stringent definition of property uniqueness (i.e. towards Method 1), the properties become more negatively selected because the regression sample will only incorporate very similar property units with multiple liens attached during the sample period.

The series in Figure D.1 include the portion of the auction bid which covers the outstanding tax debt. I repeat the exercise defined by equations (D.1) and (D.2) but using the surplus value $S_{i,t}$ as the outcome, which includes only the premium bid above the break even amount for the tax office. Such a surplus index is a closer proxy for the market value of a tax foreclosure option because it removes the portion of the face value of the lien certificate that accrues interest (see the example in Appendix A.2 for how to compute investor yields on a lien in this type of auction). For about 14% of tax sales, $S_{i,t} = 0$, indicating uncontested auctions. To accommodate zero premium bids, I apply the $\log(1 + x)$ transform, so that the new set of estimating equations is:[9]

$$\log(1 + S_{i,t}) = \delta_t + \gamma_m + \widetilde{\alpha}_i + \beta' \cdot \mathbf{X_{i,t}} + \epsilon_{i,t} \tag{D.3}$$

$$S_t = \exp(\delta_t) \tag{D.4}$$

Figure D.2 plots the resulting surplus bid index, and shows that, in contrast to the total bid index, the option value of tax foreclosure spikes in 2007, plummets during the Great Recession and does not recover. The index levels are higher for the bids made by institutional tax lien buyers, who constitute 72% of all transactions.

E Sample Local Tax Delinquency Documents

Here I provide some sample documents underlying the property tax system for Washington, D.C. Similar documents are publicly available online for other jurisdictions engaged in tax lien or tax deed sales. Washington, D.C. is somewhat unusual in that it posts PDFs of the overdue tax bill notices it sends to taxpayers, as well as revisions to the list of properties sold at tax lien auction. I scraped and combined the information across these documents to construct the set of tax sale events, and control groups of formerly delinquent but redeemed properties, analyzed in the paper. As noted in the discussion of data sources in Section 2.3, administrative tax sale records help fill in gaps in the coverage of tax auctions by transaction databases like Zillow ZTRAX and CoreLogic Involuntary Liens.

Figure E.1 displays a sample page from the ledger for revenue collections in a tax lien auction conducted by the D.C. Office of Tax and Revenue on July 17, 2018. This ledger is sometimes referred to as the "buyer's book." The buyer's book displays the square-suffix-lot (SSL, or essentially a tax assessment parcel id), address, the delinquent taxpayer on file ("To Whom Assessed"), the tax lien buyer ("To Whom Sold"), the sale date, the property tax debt ("Real Estate"), any other local tax debts ("Other Taxes"), penalties and interest accrued at the statutory 1.5% monthly rate, the starting premium bid ("Amount of Sale"), the surplus above the starting bid ("Surplus"), and the final auction price for the lien ("Total Amount"). In scenarios where the tax debt is redeemed at around the time of auction, the ledger records the redeemed amount and redemption date.

Although difficult to determine the ultimate owners of a tax-claimed property due to chains of title exchanges and repeat delinquencies, several names appear quite frequently in the buyer's book ledgers. The sample ledger page is representative of the investor composition in the D.C. tax lien market over the period 2005-2019 for which records are publicly downloadable. TIDEWATER ranks as the sixth largest institutional tax lien investor, bidding a total of $10.8 million on tax liens over this period. Another LLC listed on the page, ATCF II DC, ranks as the 15th largest institutional investor, with $5.7 million in spending. Notably, ATCF can be linked to a large asset management firm via SEC partnership agreements mentioning an ATCF special purpose vehicle, and linked to similarly named LLCs such as ATCF II Florida which are active in the Florida tax lien market.[10]

[9] The index levels are virtually identical if I instead use a inverse hyperbolic sine transform applied to $S_{i,t}$.

[10] See, for instance, the Registration Rights Agreement for Kimbell Royalty Partners, an oil and gas trust: https://www.sec.gov/Archives/edgar/data/1657788/000110465918045639/a18-16342_1ex4d1.htm.

In any contested auction, the surplus bid is strictly positive. For the transaction highlighted in red, the lienholder, TIDEWATER ASSETS LLC, bid \$78,000 above the starting premium bid relative to a total tax debt of only \$5,635.90. Figure E.2 and Figure E.3 show a final notice two weeks before tax sale. The notice was mailed to the delinquent taxpayer involved in the highlighted transaction of Figure E.1. The cover page lists the taxpayer, the SSL and address of the taxed property, the tax debt, and stipulates a payment deadline of one day before the scheduled tax sale. It then describes acceptable payment methods and goes on to list possible hardship forbearance options and income tax relief for senior citizens and low-income households (not pictured). At the end of the notice is a detachable payment slip which details the assessed value, the (rounded) mill rate of 0.85/\$100, and the annual tax bill of \$3,062.64, which implies the homestead exemption of \$75,000 applies. Combining the information between the tax notice and the lien sale, we learn that the investor paid 19.3% of the property's assessed value, or roughly 15 times the outstanding tax bill for the lien. The large surplus bid relative to the tax debt suggests TIDEWATER and competing bidders believed redemption was unlikely.

Figure E.4 excises the final page from the July 2018 buyer's book listing the total collections by tax base. For this auction nearly 100% of the collections came from liens sold against overdue property taxes. The only other source of revenue was \$58,406.81 recaptured from public utilities tax debt. Total collections from the auction amounted to roughly \$10.7 million, of which \$5.9 million came from surplus bids; the residual amount came. Only \$3.5 million, or roughly one-third of this total represents the original tax debt before interest and penalties.

FIGURE E.1. Example: Buyer's Book Page from July 2018 Tax Sale

GOVERNMENT OF THE DISTRICT OF COLUMBIA

OFFICE OF TAX AND REVENUE - OFFICE OF REAL PROPERTY TAXES

1101 4th Street, SW, Washington, DC 20024

REPORT OF THE ASSOCIATE DIRECTOR, OF THE PROPERTIES SOLD TO TAX BUYERS FOR UNPAID TAXES, SPECIAL ASSESSMENTS, WATER AND/OR SANITARY SEWER SERVICE CHARGES, AND PUBLIC SPACE RENTAL

08/22/2018

Square	Suffix	Lot	To Whom Assessed	Premise Address	To Whom Sold	Sold	Real Estate	* Other Taxes and charges Combined (see below)	Penalty, Interest, and Defective Check	Advertising	Amount of Sale	Surplus	Total Amount For Which Sold	Redemption Amount Paid	Date Redeemed Purchaser Notified, Voucher to Auditor	Special
3220		0028	M BRYANT	4513 7TH ST	TIDEWATER ASSETS, LLC RAI	07/17/2018	4,662.67	0.00	1,585.57	0.00	6,248.24	16,000.00	22,248.24			.F.
3226		0070	J REID	609 UPSHUR ST	JAMES SCHNEIDER	07/17/2018	27,374.54	0.00	12,223.77	0.00	39,598.31	54,000.00	93,598.31			.F.
3231		0121	LILLIAN SAUNDERS	4012 5TH ST	HERMAN FLORA	07/17/2018	2,097.93	0.00	589.48	0.00	2,687.41	5,500.00	8,187.41			.F.
3232		0062	CARMEN BLISSETT	533 RANDOLPH ST	TIDEWATER ASSETS, LLC RAI	07/17/2018	6,295.74	0.00	1,531.11	0.00	7,826.85	10,000.00	17,826.85			.F.
3234		0035	THE REVELY FAMILY TRUST	619 ROCK CREEK CHURCH RD	TIDEWATER ASSETS, LLC RAI	07/17/2018	4,011.06	0.00	1,183.26	0.00	5,194.32	10,000.00	15,194.32			.F.
3234		0055	CHERYL JOHNSON	517 ROCK CREEK CHURCH RD	TIDEWATER ASSETS, LLC RAI	07/17/2018	18,404.10	0.00	4,841.69	0.00	23,245.79	45,000.00	68,245.79			.F.
3237		0057	COMPASS SOLUTIONS LLC	414 SHEPHERD ST	TIDEWATER ASSETS, LLC RAI	07/17/2018	24,986.00	0.00	19,011.82	0.00	43,997.82	59,000.00	102,997.82			.F.
3260		0818	441 KENNEDY STNW ACDC LLC	441 KENNEDY ST	JUNG K. KIM	07/17/2018	3,207.76	0.00	796.49	0.00	4,004.25	3,500.00	7,504.25			.F.
3313		0048	LANETTA AGNEW	4001 MARLBORO PL	S & S INVESTMENTS LLC	07/17/2018	4,576.68	0.00	1,306.12	0.00	5,882.80	8,000.00	13,882.80			.F.
3325		0093	AJAYI DUPE	5016 2ND ST	ATCF II DC LLC, TAXSERV AS	07/17/2018	10,585.89	0.00	1,946.44	0.00	12,532.33	13,000.00	25,532.33			.F.
3364		0069	PAULA SCHRIEFER	32 WALNUT ST	TIDEWATER ASSETS, LLC RAI	07/17/2018	4,206.34	0.00	1,429.56	0.00	5,635.90	78,000.00	83,635.90			.F.
3365		0807	GLORIA LEWIS	2ND ST	JOEL F. SOP	07/17/2018	261.21	0.00	73.20	0.00	334.41	400.00	734.41			.F.

Page 24

FIGURE E.2. Example: Final Notice Cover Page for Delinquent Taxpayer

Government of the District of Columbia
Office of the Chief Financial Officer
Office of Tax and Revenue

1101 4th Street, SW
Washington, D.C. 20024

FINAL NOTICE BEFORE TAX SALE

THIS IS A NOTICE OF DELINQUENCY
FAILURE TO PAY TAXES IMMEDIATELY MAY HAVE SERIOUS CONSEQUENCES
WHICH MAY INCLUDE LOSS OF TITLE TO THE PROPERTY

Date: July 2, 2018

ALEXANDER B SCHRIEFER
PAULA G SCHRIEFER
32 WALNUT ST NW
WASHINGTON DC 20012-2153

Subject Property: SSL: 3364 0069, 32 WALNUT ST NW

TO AVOID TAX SALE YOU MUST PAY $5,635.90 by July 16, 2018.

No residence or other property improved with a building will be sold for less than $2,500, and no vacant land will be sold for less than $200.

The amount that you must pay to avoid the tax sale may be less than the total amount owed on the real property account. This amount may include fees or fines due to other DC agencies that have been certified to the Office of Tax and Revenue to be included in a tax sale pursuant to D.C. Code § 47-1340.

According to the Mayor's tax roll, you own or may have an interest in the real property listed above. Notice is given that unless you pay the amount stated above or fall within one of the limited exemptions from the tax sale, the Office of Tax and Revenue may sell this real property at tax sale.

If the property is sold at tax sale, the purchaser may have the right to file a lawsuit to foreclose on the property. You must act now to avoid additional costs and significant expenses, as well as potential loss of title to the property.

Payment to the "DC Treasurer" may be made online at *www.taxpayerservicecenter.com*, at any District branch of Wells Fargo Bank, or mailed (with the attached payment coupon) to the Office of Tax and Revenue, Real Property Tax Administration, PO Box 98095, Washington, DC 20090-8095 (please write your square, suffix and lot numbers on the check). You should keep a copy of your proof of payment in case there is a later dispute about the payment.

If payment is made less than 10 calendar days before July 16, 2018, the last business day before tax sale, you must provide a copy of the receipt directly to the Office of Tax and Revenue in order to ensure that your property is removed from the tax sale.

- You may FAX the receipt to (202) 478-5995; EMAIL the receipt to *taxsale@dc.gov*; or HAND-DELIVER a copy of the paid receipt to a Tax Sale Unit representative in the Customer Service Center located at 1101 4th Street, SW, Suite 270W, Washington, DC 20024.
- Do not mail your paid receipt.

YOU MAY BE ELIGIBLE FOR ASSISTANCE, INCLUDING A HARDSHIP FORBEARANCE OR FREE LEGAL SERVICES. PLEASE SEE THE NEXT PAGE FOR ADDITIONAL INFORMATION.

Should you have additional questions, please call OTR's Customer Service Center at (202) 727-4TAX (4829).

FIGURE E.3. Example: Final Notice Payment Stub for Delinquent Taxpayer

Payment: Payment to the "DC Treasurer" may be made online at www.taxpayerservicecenter.com or at any DC branch of Wells Fargo Bank or mailed (with payment coupon from below) to the Office of Tax and Revenue, Real Property Tax Administration, PO Box 98095, Washington DC 20090-8095 (please write your square, suffix and lot numbers on the check).

--------------------PLEASE DETACH HERE AND RETURN THIS PORTION WITH YOUR PAYMENT--------------------

Square	Suffix	Lot	Property Address	Mrtg No.	Assessment	Tax Rate/$100	Annual Tax
3364		0069	32 WALNUT ST NW	0239	$433,660	0.85	$3,062.64
ALEXANDER B SCHRIEFER, PAULA G SCHRIEFER							

Notice Number: 9370634180701
Notice Date: July 2, 2018

DCN # 1 8 1 2 0 7 1 4 8

Amount Enclosed: $ ___ , ___ , ___ . ___

For Official Use Only:

OFFICE OF TAX AND REVENUE
REAL PROPERTY TAX ADMINISTRATION
P.O. BOX 98095
WASHINGTON, DC 20090-8095

PAYMENT DUE BY:	**Jul 16, 2018**
AMOUNT DUE:	**$5,635.90**

Please ensure that the address above is visible in the window of the enclosed reply envelope

2 3364 0069 181207148 0000563590

FIGURE E.4. Example: Buyer's Book Total Proceeds from July 2018 Tax Auction

GOVERNMENT OF THE DISTRICT OF COLUMBIA

OFFICE OF TAX AND REVENUE - OFFICE OF REAL PROPERTY TAXES
1101 4th Street, SW, Washington, DC 20024

REPORT OF THE ASSOCIATE DIRECTOR, OF THE PROPERTIES SOLD TO TAX BUYERS FOR UNPAID TAXES, SPECIAL ASSESSMENTS, WATER AND/OR SANITARY SEWER SERVICE CHARGES, AND PUBLIC SPACE RENTAL

08/22/2018

Totals

Items	516		
Real Estate / Penalty and Interest on Real Estate	3,507,148.78 /	1,208,698.44	
Real Estate Special Sale	0.00		
Defective Checks	0.00		
Advertising	0.00		
* Mental Health / New York Avenue / Water and Sewer	0.00 /	0.00 /	58,406.81
* Special Assessment / Special Assessment P&I/ Bid Entities	0.00 /	0.00 /	0.00
* Public Space / Public Space Penalty and Interest	0.00 /	0.00	
* Clean City / Clean City Penalty and Interest	0.00 /	0.00	
* DCRA / DCRA Penalty and Interest	0.00 /	0.00	
* Vacant and Abandoned / Vacant and Abandoned Penalty and Interest	0.00 /	0.00	
* MTB / MTB Penalty and Interest	0.00 /	0.00	
* Title Search Fee / Recording Fee	0.00 /	0.00	
Surplus	5,890,600.00		
Buyers Book Grand Total			**10,664,854.03**

F Linking Private Equity Deals to Tax Sale Records

Table F.1. Private Equity Deals Involving Formerly Tax-Distressed Properties

Deal date	Buyer(s)	Sellers(s)	Property name	Current use	Deal size	Square footage	Tax sale date	Auction price	Lien buyer(s)
1/2010	The Goldstar Group	Unidentified	The Floridian	Condominiums	N/A	N/A	12/2009	$4,472	CAPITAL SOURCE BANK FBO
2/2010	CBRE Investment Management	PGIM Real Estate	Mass Court†	Multi-family	$100.5 M	200,000	9/2008	$35,005	US BANK-CUST/SASS MUNI V DTR
4/2010	Somerset Development Co. THC Affordable Housing Unidentified buyer(s)	Unidentified	Webster Gardens Apartments	Multi-family	N/A	N/A	7/2006	$7,768	CAPITOL TAX SERVICES, LLC
7/2010	Cadila Healthcare Ltd.	Akridge	1016 16th St	Mixed-use	N/A	38,250	11/2009	$8,548	ETS DC LLC
3/2011	Wells Real Estate Funds	Beacon Capital Partners	Market Square	Mixed-use	$615 M	681,051	11/2009	$5,309	SALTA GROUP INC
6/2011	JCR Companies	Unidentified	301 Massachusetts Ave NW	Retail	N/A	4,000	9/2008	$11,986	REDEMPTOR LITIUM LLC
7/2011	Greystar Real Estate Partners	Lehrco	The Warwick†	Multi-family	$65.1 M	460,262	9/2010	$39,086	MUNITRUST CAPITAL FUND II LLC HEARTWOOD 20, LLC
8/2011	JCR Companies	Unidentified	1723 Connecticut Ave NW	Residential	N/A	N/A	7/2011	$52,144	US BANK CUST/EMP IV, CAP ONE
7/2012	Host Hotels & Resorts	Quadrangle Development Corporation	Grand Hyatt Washington	Hotel	$400 M	N/A	9/2010	$9,505	REDEMPTOR LITIUM LLC
6/2013	The Goldstar Group	Unidentified	425-427 8th St SE	Mixed-use	$1.7 M	3,155	7/2012	$23,880	REDEMPTOR LITIUM
9/2013	Rose Green Cities Fund	Unidentified	Portner Flats	Multi-family	N/A	40,860	7/2006	$15,596	HEARTWOOD 88, LLC
11/2013	Lone Star Funds	Wereldhave	1401 New York Ave NW	Office	$71 M	300,000	12/2010	$371,957	WCP DC23 HOLDINGS, LLC
11/2013	Urban Investment Partners Perseus TDC PRP	Unidentified	Capitol Park Tower	Multi-family	N/A	N/A	9/2008	$8,507	REDEMPTOR LITIUM LLC
12/2013	Hines	Lehman Brothers Real Estate Private Equity MacFarlane Partners Monument Realty	55 M St	Mixed-use	$141.9 M	266,566	7/2006	$8,661	AEON PROPERTIES, LLC
6/2014	Rezross Investment Group LLC	DelShah Capital	2209 Massachusetts Ave NW	Office	N/A	9,000	7/2012	$50,494	REO AMERICA INC CUST AS
7/2014	Jamestown	Vornado Realty Trust Angelo Gordon & Co.	Georgetown Park	Mixed-use	$319 M	315,028	9/2008	$15,036	AEON FINANCIAL IV, LLC - SERIES
12/2014	PRP	Unidentified	2501 Residences on Street	Condominiums	$31.6 M	100,000	9/2011	$4,873	RICHARD T. COCKERILL
6/2015	Mill Creek Residential	Potomac Construction Group The Garrett Cos. Valor Development	2700 16th St	Office	$16.2 M	35,000	7/2012	$160,842	ABBOTT DEVELOPMENT GROUP
10/2015	Blackstone Group	Columbia Property Trust	Market Square	Mixed-use	$291.6 M	686,000	7/2012	$5,542	ELM CAPITAL LLC
4/2016	JCR Companies	Unidentified	916 G St NW	Retail	$3.2 M	6,200	9/2008	$11,557	POTOMAC TAX SERVICES LLC
4/2016	KHP Capital Partners	Xenia Hotels & Resorts	The Darcy Hotel	Hotel	$65 M	N/A	12/2010	$56,435	WCP DC23 HOLDINGS, LLC
11/2016	NOVO Properties	Unidentified	1440 Chapin Street Apartments	Multi-family	$4.1 M	N/A	7/2007	$6,935	US BANK-CUST/SASS MUNI V DTR
4/2017	Pearlmark Real Estate Partners Lincoln Property Company	Unidentified	619 14th St NW	Office	$57 M	35,000	9/2008	$12,797	TCF NATIONAL BANK FBO AEON
6/2017	Jefferson Apartment Group Stars Investments	Bristol Capital Corporation Multi-family	2009 8th St NW	Multi-family	$25 M	25,050	7/2005	$50,597	HEARTWOOD 88, LLC
1/2018	Atalaya Capital Management Urban Investment Partners	JP Morgan Asset Management	Onyx on First	Multi-family	$95.5 M	N/A	7/2006	$17,084	MUNITAX FUND, LLC
3/2018	TA Realty	JP Morgan Asset Management	Gables Woodley Park	Multi-family	$106.5 M	N/A	7/2016	$23,282	DUPONT TAX SERVICES, LLC
8/2018	Marx Realty	Unidentified	819 7th St NW†	Mixed-use	$11.6 M	21,952	7/2016	$229,421	SUNSHINE STATE CERTIFICATES NEBRASKA ALLIANCE REALTY
8/2018	Artemis Real Estate Partners Jair Lynch	Level 2 Development FCP	Takoma Central	Multi-family	$50.6 M	82,935	7/2007	$27,676	FEDOR TOMENKO
10/2018	Aria Investment Partners	Unidentified	23 Florida Ave NE	Development land	N/A	25,000	7/2012	$66,682	ABBOTT DEVELOPMENT GROUP
1/2019	Kodiak Properties	Unidentified	1539 7th St NW	Mixed-use	$2.2 M	3,000	9/2010	$7,798	HMTR I, LLC
3/2019	Next Realty	Unidentified	Metropole	Condominiums	$4.8 M	N/A	7/2016	$18,934	NEBRASKA ALLIANCE REALTY

Notes: Deals listed in the table hand-matched between public tax sale records and Preqin single-asset real estate deals conducted between 2005 and 2019 for properties in Washington, D.C. I order the deals chronologically by deal date. Deal size recorded in millions of USD. N/A refers to missing values in the Preqin database. Auction price refers to the final bid at the tax sale auction (outstanding tax debt + surplus bid). The † symbol indicates properties for which liens were placed on multiple lots within the parcel prior to the deal acquisition date. In such cases, I total the auction prices across the liens. This list includes only deals resulting in a tax foreclosure, and information attached to the last recorded tax lien sold for that property. See Section 2.2 for more details on sample construction.

Appendix References

Butts, K. (2021a): "Difference-in-Differences Estimation with Spatial Spillovers," *mimeo*, University of Colorado Boulder.

Butts, K. (2022): "Difference-in-Differences with Geocoded Microdata," forthcoming, *Journal of Urban Economics: Insight*.

Caliendo, L., F. Parro, E. Rossi-Hansberg, & P-D. Sarte (2018): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *Review of Economic Studies*, 85(4): 2042-2096.

Chetty, R., J. Friedman, N. Hendren, M.R. Jones, & S.R. Porter (2018): "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," NBER Working Paper, No. 25147.

Diamond, R. & T. McQuade (2019): "Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Development," *Journal Political Economy*, 127(3): 1063-1117.

Fang, H., Q. Gu, W. Xiong, & L. Zhou (2015): "Demystifying the Chinese Housing Boom," NBER Macroeconomics Annual, 30: 105-166.

Ganduri, R. & G. Maturana (2021): "Do Property Rehabs Affect Neighboring Property Prices?" *mimeo*, Emory.

Gaubert, C., P.M. Kline, D. Vergara, & D. Yagan (2021): "Trends in U.S. Spatial Inequality: Concentrating Affluence and a Democratization of Poverty," NBER Working Paper, No. 28385.

Loftis, L.B. (2007): *Profit by Investing in Real Estate Tax Liens*, Second Edition, Kaplan: New York, NY.

McMillen, D.P. (2012): "Repeat Sales as a Matching Estimator," *Real Estate Economics*, 40(4): 745-773.

Owens, R., E. Rossi-Hansberg, & P-D. Sarte (2020): "Rethinking Detroit," *American Economic Journal: Economic Policy*, 12(2): 258-305.

Schmidheiny, K. & S. Siegloch (2020): "On Event Studies and Distributed-Lags in Two-Way Fixed Effects Models: Identification, Equivalence, and Generalization," CEPR Discussion Paper, No. 13477.