THE UNIVERSITY OF CHICAGO

PROPERTY TAX AND RESIDENTIAL CHOICES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
MENGYUE HE

CHICAGO, ILLINOIS
JUNE 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am immensely grateful to my dissertation committee, Ali Hortaçsu, Michael Dinerstein, and Christopher Berry for their continued guidance and support, not only during my thesis writing but also throughout my whole graduate school.

I thank Infogroup, Zillow, CoreLogic, and Costar for providing data for this study. I am also thankful to the Research Computing Center for providing data and technical support.

I would like to thank my friends and peers, as well as participants in the Industrial Organization workshop for their feedback and discussion.

I am thankful to my parents, who always support me with unconditional love and cover me with prayers.

I cannot thank enough to my husband, Ce, for his love, support, and encouragement. I am so grateful that we can go on the life journey together.

The past years have taught me a lot about "A man's heart deviseth his way; but the LORD directeth his steps." All honor and glory to God.

ABSTRACT

Due to outdated property assessment systems, property assessment does not match with transaction price in most cases. Assessment is used as a measurement of a property's market price and property tax is a fixed portion of the assessment. The effective tax rate (rate of a property tax over a property's market price), as a result, is unequal within jurisdictions. Philadelphia was one of those jurisdictions before 2014 and then improved substantially through the Actual Value Initiative in 2014. Using data on households' residential address history and real estate tax and transaction records, I first show that property tax increases made homeowners more likely to move out and less likely to keep owning their properties. Then I construct a model characterizing households' location and homeownership choices in equilibrium and estimate the model. The counterfactual analysis shows that if we correct the current unequal property tax system and change it to a flat-rate one, households will be more likely to be homeowners, especially low-income and less wealthy households. The welfare analysis shows such a change will increase the overall households' welfare by 3.53%. Specifically low and middle-income households enjoy higher welfare gain while high-income households will get a slight welfare reduction. The government will get 23.64% more property tax income.

CHAPTER 1

PROPERTY TAX AND RESIDENTIAL CHOICES

1.1 Introduction

Property tax is the largest source of revenue for local governments in the US. In 2017, local governments collected nearly half of their own-source general revenue from property tax. On the other hand, property tax is an important part of housing expenses for homeowners. Nationwide, the effective property tax rate is 1.1% of the average home value. The property tax payment is about 30% as much as the mortgage payment for the homeowner (at an interest rate 3%). Despite the importance of property tax, few literatures study how property tax affects social welfare by changing households' choices.

Property tax is designed to be *ad valorem* tax, meaning the tax amount should be proportional to the property's value. In reality, property value is measured through assessment, so property tax equals assessment value times tax rate. Berry (2021) found that in nearly 97.7% of counties throughout the US, the ratio of property's assessment value to its actual sale price for the bottom 10 percent priced properties is greater than the ratio for the top 10 percent priced properties. As a result, the rate of property tax bills over property price is regressive. Many reasons could lead to such regressivity, including outdated assessment, data limitation, conditional averaging models, and assessment growth cap.

Such a non-proportional property tax system makes households pay at different effective tax rates (tax amount/market value), so it disproportionately burdens households. The welfare impact on each household is unequal. When we aggregate up to social welfare, it is even more ambiguous what efficiency impact such a non-proportional property tax system has. In economics, we know a perfectly proportional tax system could cause welfare loss because of price distortion. A non-proportional tax system could further make the efficiency loss higher or lower.

To figure out how such a non-proportional property tax system could affect welfare, we will need to first understand what decisions households make in the residential market. When households make location choices, they not only choose where to live but also choose whether to rent or buy a property. Similarly, when households choose whether to rent or buy a property, they will also face the location choice. Because these two types of choices are usually bonded together, policies affecting one type of choice could have very complicated impacts because people change the other type of choice. Property tax policy, by affecting the cost of owning properties, could make households relocate. Their relocation choices furthermore could bring heterogeneous influences across regions and population groups.

This study means to bring the two types of choices together and understand how these two dimensions are affected by property tax. The objective is to measure the welfare influences of inequitable property tax systems compared with a flat-rate property tax system.

Distinguishing renters and owners in this analysis is important in several ways. First, the policy itself affects homeowners and renters differently. Owners' cost is directly affected by property tax, while renters may be indirectly affected by market supply and demand change caused by relocation as well as the resulted rent change. Second, renters and owners could have different preferences on location choice, which results in unequal distribution across regions. Regions with higher homeownership rates could be more affected. The different preferences will also affect where owners and renters move to while relocating, which could cause influences dis-proportionally distributed across regions. Third, the policy could affect people changing their ownership choice.

Empirically, I use a policy reform in Philadelphia in 2014 which created an exogenous shock on property tax by reevaluating all properties. The goal is to answer two questions: Does property tax increase affect households' migration probability? Does property tax increase affect households' homeownership decisions? When property tax adjusts, homeowners face a direct impact from the increasing housing cost. The adjustment of rent, as a result of reallocation, usually comes later. How households respond to the property tax change varies by whether they were a homeowner or renter before the change. Therefore, I analyzed the two questions separately for homeowners and renters.

To answer these questions, I use a novel dataset from Infogroup Consumer Database, a commercial database that tracks the residential address history of US households from 2006. I combine this dataset with ZTRAX and CoreLogic datasets where I get property tax and ownership information. I construct a sample that includes all households who lived in Philadelphia in 2013.

Using the merged dataset, I evaluate how households' migration decisions and homeownership decisions are affected by property tax increases. I grouped households to homeowners and renters by their homeownership status in 2013 and evaluate the impact separately for the two groups. The difference-in-differences results show that 1000$ more property tax increased the probability that homeowners moved out by 0.28 percentage points, which was 15% increase relative to the average migration rate. The property tax increase also made renters more likely to move-out, but at a much lower magnitude and a bigger p-value (0.1 vs. 0.01). 1000$ increase in property tax increased the probability that an owner became a renter by 0.19 percentage point, which is 5% of all households that have ever been renters in the sample period.

In order to evaluate the welfare impact of property tax systems, I construct an equilibrium model characterizing households' joint decisions of location and homeownership. I first estimate the baseline model by methods of moments. The estimation results show that high-wealth households prefer owning and low-wealth households prefer renting. The magnitude of utility on renting of low-wealth households and the disutility on renting of high-wealth households are very close. The utility of log consumption is much larger. For a high-wealth household with income $100,000 and $10,000 annual housing payment, conditional on housing payment, the extra utility of owning a property compared with renting provides the same utility gain as having $39,000 money. I also check the robustness of the estimation to heterogeneous preference on the neighborhood, unobserved heterogeneity, and the ratio of annual housing payment to house price.

Finally, because property tax is designed to be a flat rate, I perform a counterfactual exercise to study how changing the property tax system to flat-rate affects households ownership choices. I find that such change will increase the homeownership rate of low-income and low-wealth households by around 4%. The welfare analysis shows that changing the property tax system to flat-rate will increase the welfare of low and middle-income households by 3% or more. High-income households will get a welfare reduction of less than 0.3%. The overall population welfare gain is 3.53%. The government will get 23.64% more property tax income.

Very few economic literatures have empirically measured how property tax policies affect mobility. Ding & Hwang (2020) studied the AVI policy and find that gentrifying neighborhoods, compared with non-gentrifying ones, on average experienced $540 increase in their annual tax amount and 4.1 percentage point increase in the tax delinquency rate post-AVI, with the largest increases happening to neighborhoods that underwent intense gentrification, which were $1045 and 6.1 percentage points, respectively. Using aggregate migration data from a consumer survey, they also showed that within five years after the adoption of the AVI, the probability of moving for elderly or long-term homeowners did not change significantly. Shan (2010) uses variation in state property tax policy as an identification strategy and finds that higher property taxes significantly raise the mobility of elderly homeowners. My study will be able to leverage more detailed micro-level migration data and also compare owners and renters.

As for the residential choice model, few literatures have ever incorporated the rental and purchase markets at the same time in residential sorting models. Some only use the real estate transaction data to estimate the model (Bayer et al. (2016)). Others usually just consider the equilibrium in the rental market (Almagro & Domínguez-Iino (2021)). Binner & Day (2015) present an equilibrium sorting model with simultaneous rental and purchase markets in which tenure choice (choice on renting or purchasing a property) is endogenized, but they only calibrate the model with limited aggregate data.

The rest of this paper is organized in the following way. Section 2 of this paper provides an overview of the institutional background. Section 3 introduces the data used in this paper. Section 4 describes the empirical strategy I use to evaluate the policy impact and presents the results. Section 5 presents an equilibrium model characterizing the joint decisions in the housing market and the estimation strategy. Section 6 discusses the estimation result and counterfactual analysis. Section 7 concludes.

1.2 Institutional Background

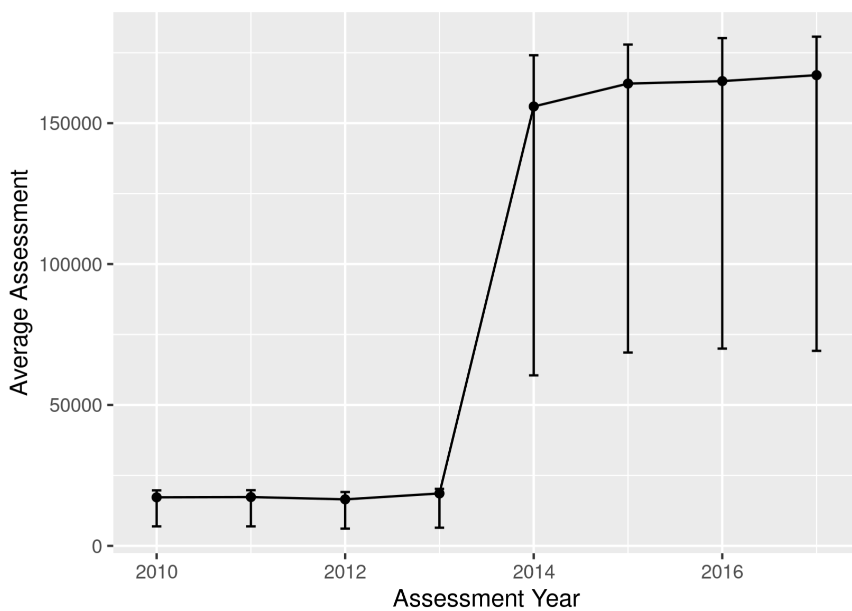
Actual Value Initiative (AVI), a major overhaul of the property tax system, in Philadelphia, was adopted in 2013 and took effect in 2014. Until 2013, there had not been a comprehensive reassessment of properties' market values in Philadelphia since the 1980s. Thus, the assessment values for most properties had largely been unchanged for several decades and so deviated from the actual market values. In 2014, the city reassessed all properties, trying to make assessment values closer to properties' actual market values. As a result, the AVI generated significant variations in property taxes across properties.

Before 2013, the city levied taxes on 32 percent of a property's assessment value. Take 2013 as an example when the tax rate was 9.771. For a property with assessment value $300,000, the tax bill would have been $9390 ($\$300,000 \times 0.32 \times 0.09771$). Effective tax rate, defined by the ratio of tax amount to market value, would have been 3.13. After the AVI, 100 percent of properties' assessment value is used to calculate tax bills. The tax rate was also adjusted from 9.771 percent in 2013 to 1.34 percent in 2014 to keep the adjustment revenue-neutral. Using the previous example, a property assessed at $300,000 in 2014 would have been taxed \$4020 ($\$300,000 \times 0.0134$). Thus, properties experienced an increase in tax amount if the assessment value increased by more than 6.29. The AVI generated significant variations in property taxes across properties, which I will use to evaluate how households respond to property tax changes.

I used CoreLogic datasets to get the total assessment amount and total property tax amount information of properties in Philadelphia county and summarize the changes over time.

Figure 1.1 shows the average assessment value of Philadelphia properties with 25% and 75% percentiles from 2010 to 2017. There is a clear jump in 2014 which reflects the AVI policy influence.

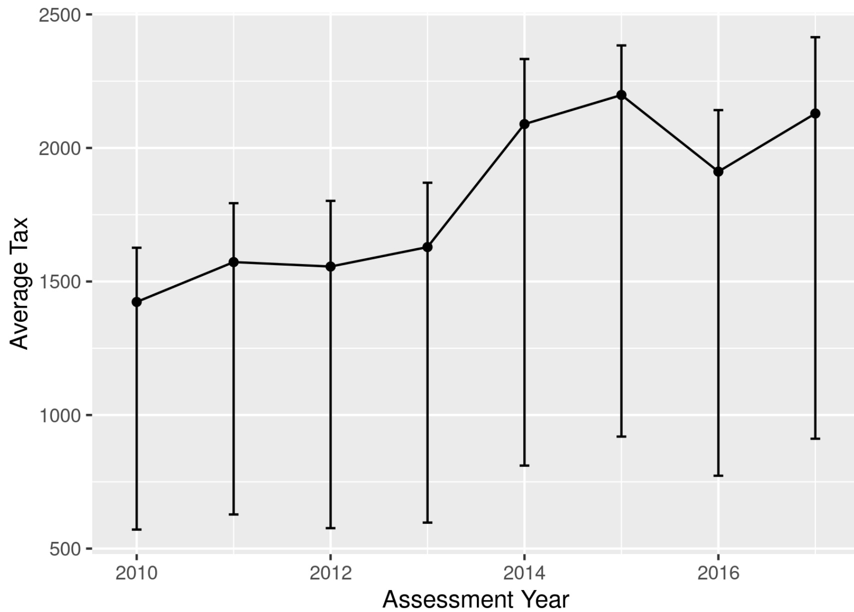
Figure 1.1: Total Assessment Value of Philadelphia Properties



Note: Figure shows the distribution of assessment value of Philadelphia properties from 2010 to 2017. The points are the mean. The bars are the 25% and 75% percentiles.

Figure 1.2 shows the average tax amount of Philadelphia properties with 25% and 75% percentiles from 2010 to 2017. Property tax is not only affected by assessment value, so the trend is a bit different. But from 2013 to 2014, the average tax increased almost \$500 (30%).

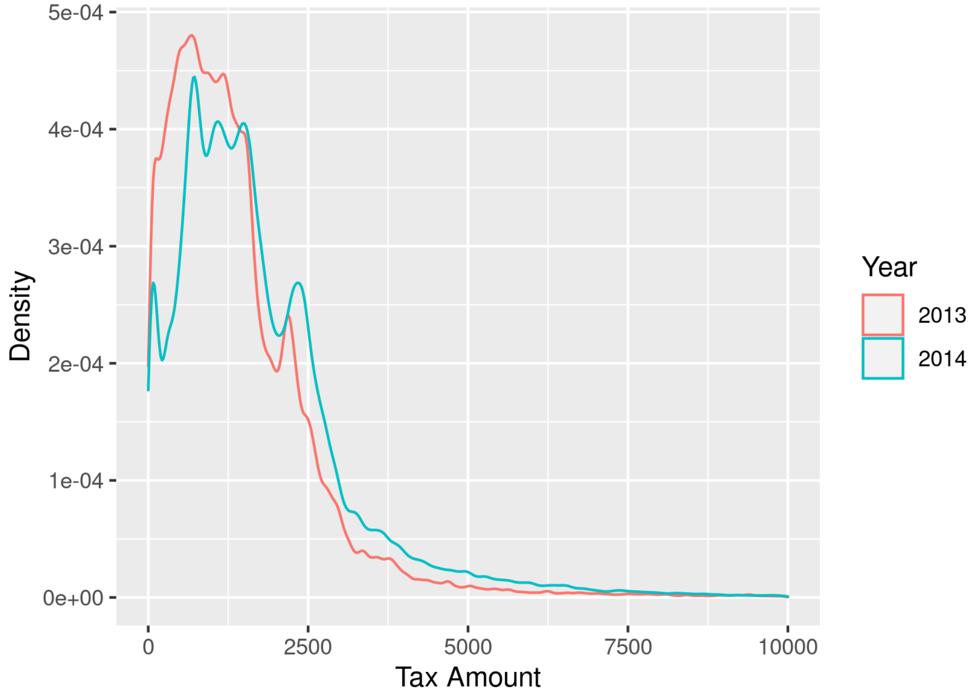
Figure 1.2: Property Tax Amount of Philadelphia Properties



Notes: Figure shows the distribution of total property tax amount of Philadelphia properties from 2010 to 2017. The points are the mean. The bars are the 25% and 75% percentiles.

Figure 1.3 shows the total tax amount distributions of properties in Philadelphia. The distribution moved to the right in 2014.

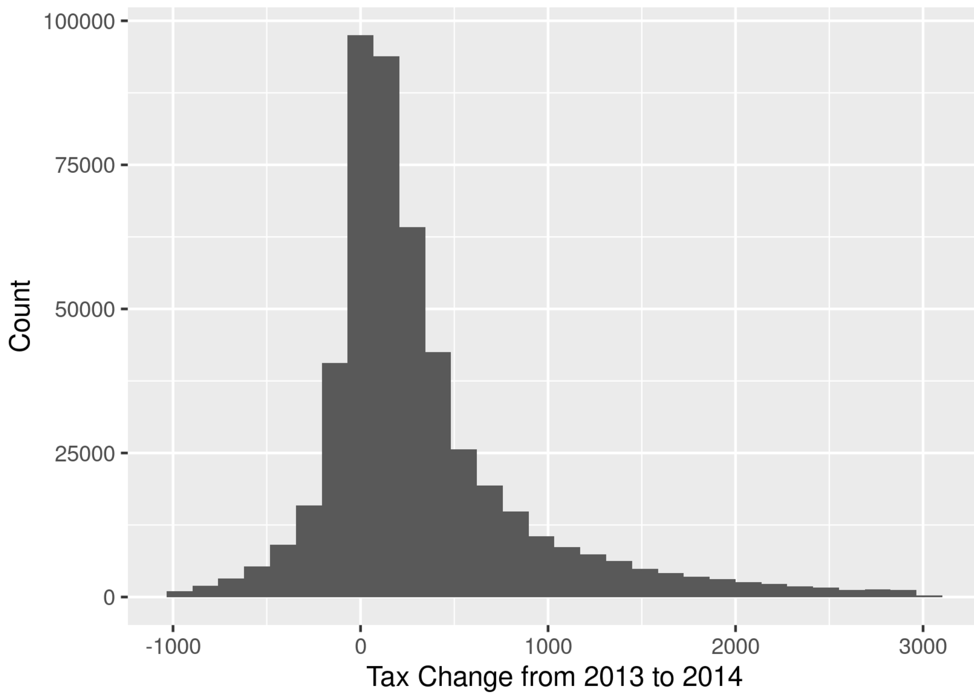
Figure 1.3: Tax Amount Distributions in 2013 and 2014



Notes: Figure shows the distribution density of total property tax amount of Philadelphia properties in 2013 and 2014.

Figure 1.4 shows the histogram of property tax increase of Philadelphia properties.

Figure 1.4: Property Tax Increase in Philadelphia from 2013 to 2014



Notes: Figure is the histogram of the property tax difference for properties in Philadelphia between 2013 and 2014 (tax in 2014 minus tax in 2013).

Figure A.1 in the Appendix shows the geographical variation of the tax change by Census tracts. Figure A.2 shows the median income of each Census tract. Through comparing the two figures, we can find that the tax change does not obviously tie with median income. I also test some other neighborhood characteristics and find that it does not correlate with median education level or population composition.

1.3 Data

The data used in this paper come from multiple sources. I will briefly describe the data processing procedure in this section and leave the details in the appendix.

1.3.1 Data Source

Infogroup

Infogroup Data provide residence information for US households for the years 2006-2017, with national coverage. Each observation is a household in a year, which is a snapshot of the address and some household characteristics. For each household, the data record up to three household members and their names. The dataset also has variables describing household characteristics, including household income, household wealth, and an indicator for children.

For each address in the dataset, there is a variable showing the month and year of confirmation. If a household is observed in 2017 with a confirmation date in 2016, I create an observation of this household in 2016 by copying the information in 2017. In this way, I fill in the gaps in the address history.

CoreLogic

CoreLogic tax data contain the assessment and tax information of properties across the US from 2009 to 2017. I extract the total assessment amount and total property tax amount information of properties in Philadelphia county and drop the data in 2009 because of poor quality.

ZTRAX

Zillow's Transaction and Assessment Database (ZTRAX) is a dataset provided by Zillow Group, Inc. The dataset has two parts. One is the assessment dataset and the other is the transaction dataset. The assessment data has property addresses and some construction details. The transaction dataset covers deed transfer, mortgage, and foreclosure records. For each deed transfer, it provides the buyer and seller's names.

Using ZTRAX assessment data, I create a panel dataset on property-assessment year level showing the address and building characteristics of all properties in Philadelphia.

Then I create a panel dataset of Philadelphia properties on property-year level recording all owners' last names of the property in each year using ZTRAX transaction records and supplement it with the ownership information in the assessment data in 2018.

Costar

Costar dataset provides rent of multi-family buildings across the years. Using that dataset, I get the average rent for each zip code area in each year.

1.3.2 Data Match

I first merge the CoreLogic tax data to ZTRAX assessment data using Assessment Parcel Number, the unique parcel identifier from the county's assessment office, and assessment year. Then I merge the ownership data to the tax and assessment data using Assessment Parcel Number and year. If an observation in the tax and assessment data matches with multiple ownership observations, I combine the owner's last names from the multiple matches. After these two steps, I get a panel dataset on the property-year level. Finally, I merge the property panel data to Infogroup data using addresses. Appendix C.4 has more details on the data merge process.

To evaluate the match quality, I summarize the number of households living in Philadelphia in the Infogroup data and the percentage that could find matches in the property data in Table 1.1. The match rate is pretty high and stable over time. Around 70% of the mismatches are due to data quality issues. The other 30% are likely to be commercial buildings where the whole building is taken as a single unit for assessment and tax purposes. The match rate is higher in more populated, lower-income, less educated, and fewer white people proportion regions. More details are in C.5.

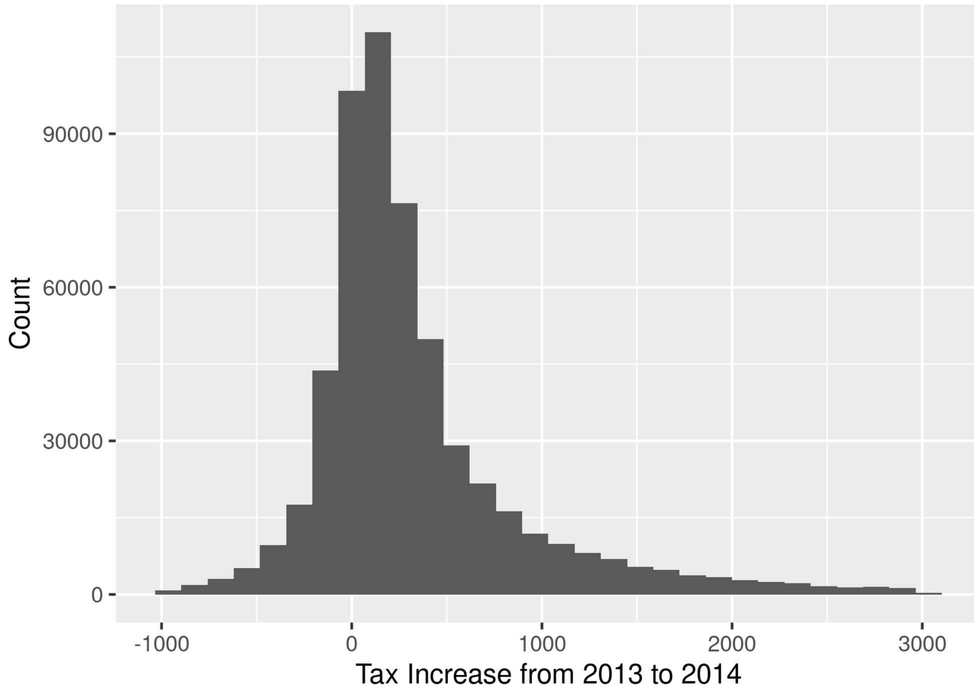
Table 1.1: Infogroup Philadelphia Household Count and Match Rate

Year	Household Count	Matched Households	Match Rate
2006	788652	619811	78.6%
2007	788684	618988	78.5%
2008	671813	528808	78.7%
2009	740959	589151	79.5%
2010	788091	624580	79.3%
2011	825571	654644	79.3%
2012	899164	709104	78.9%
2013	897711	704556	78.5%
2014	880484	693917	78.8%
2015	943099	745582	79.1%
2016	913886	728420	79.7%
2017	862818	692922	80.3%

Notes: Table shows the number of households living in Philadelphia in the Infogroup data, the number and the percentage of those observations that could find matches in the property data. The match is Property dataset is constructed by merging the CoreLogic tax dataset and ZTRAX dataset.

With the merged dataset, Figure 1.5 shows the histogram of property tax increase for households living in Philadelphia in 2013. On average, households got a property tax bill increase by $521.45. This plot and 1.4 have very similar distributions, which also shows how representative the Infogroup household database is.

Figure 1.5: Philadelphia Household Property Tax Increase from 2013 to 2014



Notes: Figure is the histogram of the property tax difference for households between 2013 and 2014 (the property tax amount for one household in 2014 minus the property tax amount for the same household in 2013). The sample is households living in Philadelphia in 2013.

Appendix C.5 has more discussions on the validity of the Infogroup dataset.

1.4 Policy Effect on Moving and Homeownership

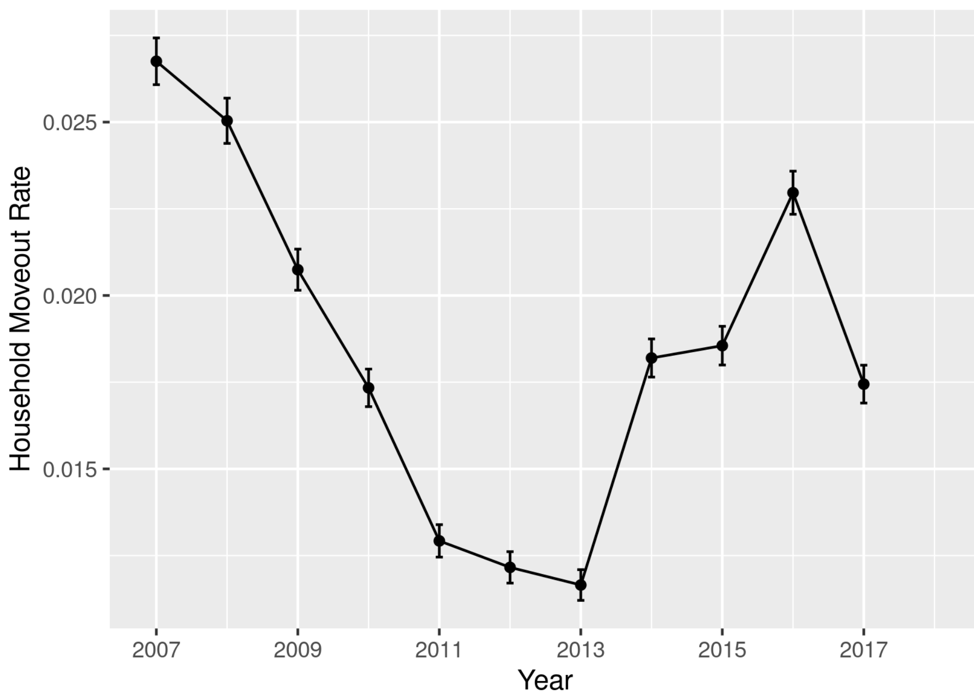
1.4.1 Empirical Strategy

Using the household-year level data, I construct two outcome variables. One is the ownership indicator which takes value 1 if any individual in the household has a last name that is in the owners' last name list. The other is the move indicator. If a household lives at an address with a different house number, street number, or unit number in the next year, I define the move-out indicator to be 1.

Among all the households in the dataset, I took out the households that have observations in all the 12-year time span and summarize their migration and homeownership patterns across the years.

Figure 1.6 shows the average move-out rate of all households in the balanced panel. The bars are the 95% confidence intervals. The figure shows that the average migration rate was dropping before 2013, but increased in 2014 and the following years.

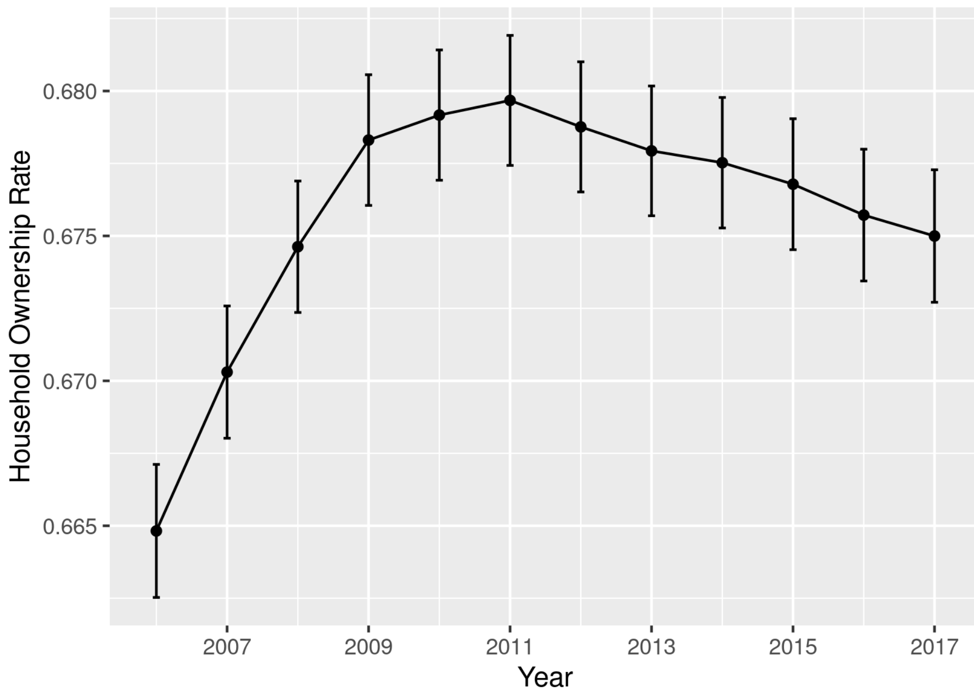
Figure 1.6: Average Migration Rate



Notes: The sample includes households living in Philadelphia in 2013 and having observations in Infogroup data from 2006 to 2017. Figure shows the average move-out rate of all households in the balanced panel. The points are the average rate. The bars are the 95% confidence intervals.

Figure 1.7 shows the average ownership rate of all households in the balanced panel. The average ownership rate did not change much before and after the AVI reform.

Figure 1.7: Average Homeownership Rate



Notes: The sample includes households living in Philadelphia in 2013 and having observations in Infogroup data from 2006 to 2017. Figure shows the average ownership rate of all households in the balanced panel. The points are the average rate. The bars are the 95% confidence intervals.

No obvious ownership rate change, however, does not mean there are no individual households changing their ownership decision because of the reform. To evaluate how property tax increases could affect households' choices, I use a differences-in-differences model to evaluate whether households experiencing bigger tax increases in 2013 were more likely to own the property or move-out.

$$Y_{hpt} = \alpha Z_h + \beta Z_h \mathbb{1}\{t \geq 2013\} + \eta_p + \delta_t + \epsilon_{hpt} \tag{1.1}$$

The sample contains all households living in Philadelphia in 2013. The treatment variable, Z_h, is on the household level, measuring the tax increase from 2013 to 2014 of the property where household h lives in 2013. I control property fixed effect η_p, and year fixed effect δ_t. All standard errors are clustered on the household level.

I also did some robust checks by adding property-level fixed effect and clustering the standard errors on both the household level and property level. Results are shown in the Appendix.

Because moving usually has underlying cycles, a household fixed effect may not be good at controlling such cycles. Instead, a property fixed effect could capture the cyclical influence because households living in the same apartment or single-family house might share a common moving pattern whereas one household living in different types of properties at different points of a life-cycle can have very different moving patterns. Therefore, I control property fixed effect instead of household fixed effect in the model. The results with all three fixed effects are very similar.

I also estimate the model separately for households that were owners in 2013 and households that were renters in 2013.

The following event-study regression helps check the no pre-trend assumption:

$$Y_{hpt} = \alpha Z_h + \sum_{\tau=t_{min}}^{\tau=t_{max}} \beta_\tau Z_h \mathbb{1}\{t = \tau\} + \eta_p + \delta_t + \epsilon_{hpt} \tag{1.2}$$

Because the AVI policy created property tax variation by reassessing all the properties, how much property tax increase a household got was correlated with how long the property had not been assessed and the price increase before AVI. They both did not vary before and after AVI. So, there are no important time-varying confounders that correlate with tax increase and may also affect the outcome. Using the difference-in-differences approach is valid in this case.

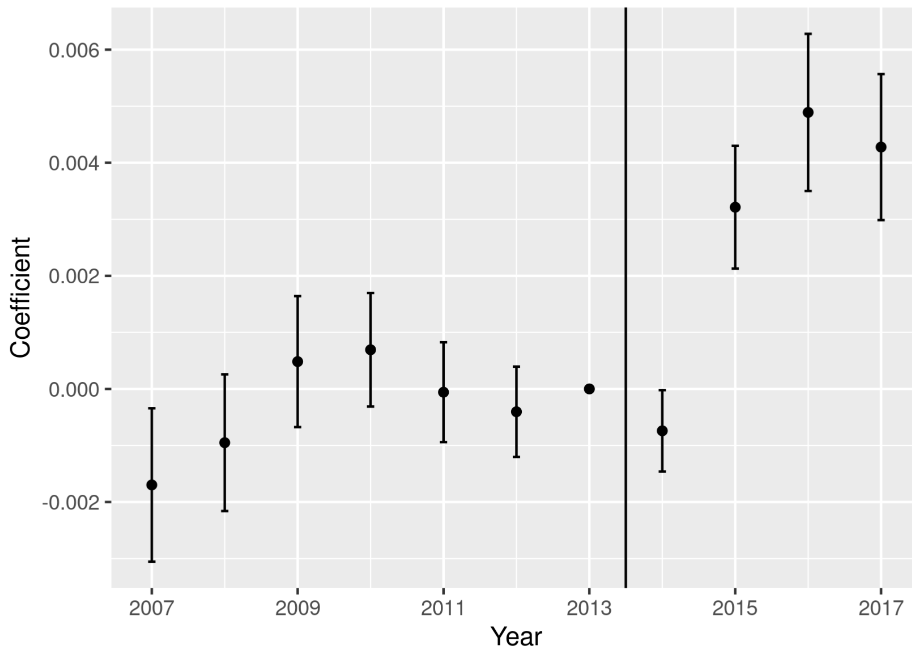
1.4.2 Moving

Figure 1.8 shows the event study plots using moving out as the outcome variable. The sub-figure (a) is the plot for owners and (b) is for renters. The y-axis is the coefficient β_T. The x-axis is the year. In each year, there is a point estimate and a 95% confidence interval shown by the bar. I normalize the coefficient in 2013 to zero.

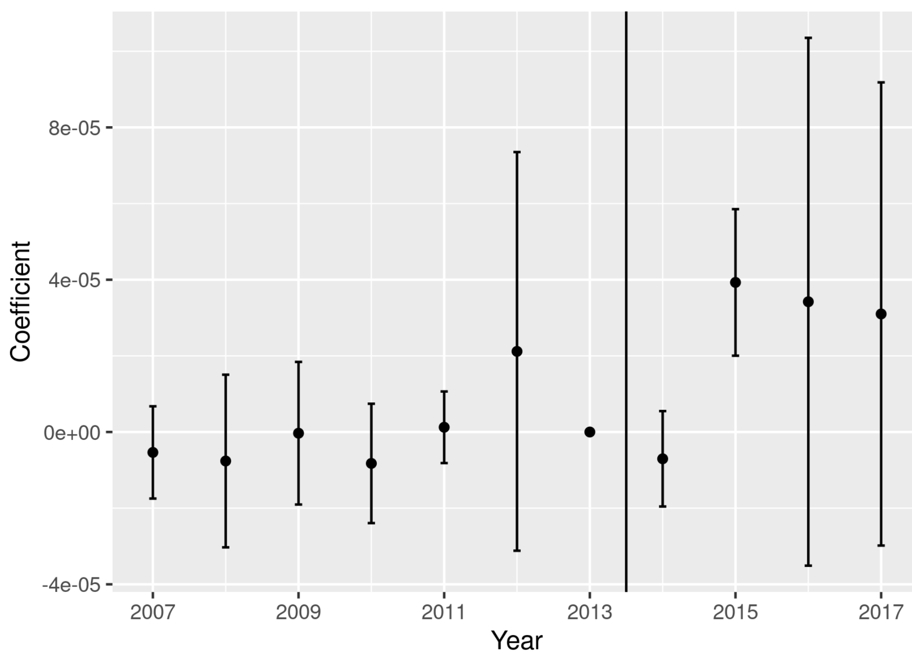
The figure shows no pre-trend before the reform both for owners and renters. For owners, the move-out rate increased with property tax increases since 2015. For renters, the move-out rate increased a little with property tax increases in 2015, but not in 2016 or 2017.

Figure 1.8: Move-out Event Study

(a) Owners



(b) Renters



Notes: Figure shows estimated coefficients (and 95% confidence intervals) on property tax increase ($1,000s) for each year from the difference-in-difference regression of moving indicator on the property tax increase, property fixed effects, and year fixed effects. The 2013 coefficient is normalized to 0.

The difference-in-differences regression results are shown in Table 1.2. Column (1) shows the results for the whole sample. Columns (2) and (3) show the results for the owner sub-sample and renter sub-sample. The table shows that 1000$ more property tax increased the probability that owners moved out by 0.28 percentage points, which was 15% increase relative to the average migration rate. It also increased the renter's migration probability, but with a much smaller magnitude and significance. Because the tax increase is household level and fixed effects are property level, I also control tax increase in the estimation. The results show that households with higher tax increases in 2014 have lower average migration rates. But homeowners with higher tax increases in 2014 on average have 1.17% higher migration rates.

Table 1.2: Household Level Dif-in-Dif Results

Dependent Variable:		Move-out	
	All	Owners	Renters
Model:	(1)	(2)	(3)
Tax Increase	-0.0002**	0.0117**	-8.383×10^{-5}
	(6.827×10^{-5})	(0.0047)	(6.93×10^{-5})
Tax Increase × I(YEAR≥2013)	2.236×10^{-5}**	0.0028***	1.244×10^{-5}*
	(1.131×10^{-5})	(0.0004)	(6.575×10^{-6})
Parcel FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	4,327,257	2,263,190	2,051,542
R^2	0.20274	0.26661	0.26272
Mean of Dependent Var	0.0196	0.0190	0.0381

Notes: Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. Data span 2007 through 2017. fixed effects. Standard errors are clustered by property. "Tax Increase" is the property tax increase (in $1000s) from 2013 to 2014 of the property that the household lives at in 2013. "I(YEAR≥2013)" is the post AVI reform indicator. Column (1) uses the full sample, which includes all the households living in Philadelphia in 2013. Column (2) includes households living in Philadelphia in 2013 and were homeowners. Column (3) includes households living in Philadelphia in 2013 and were renters.

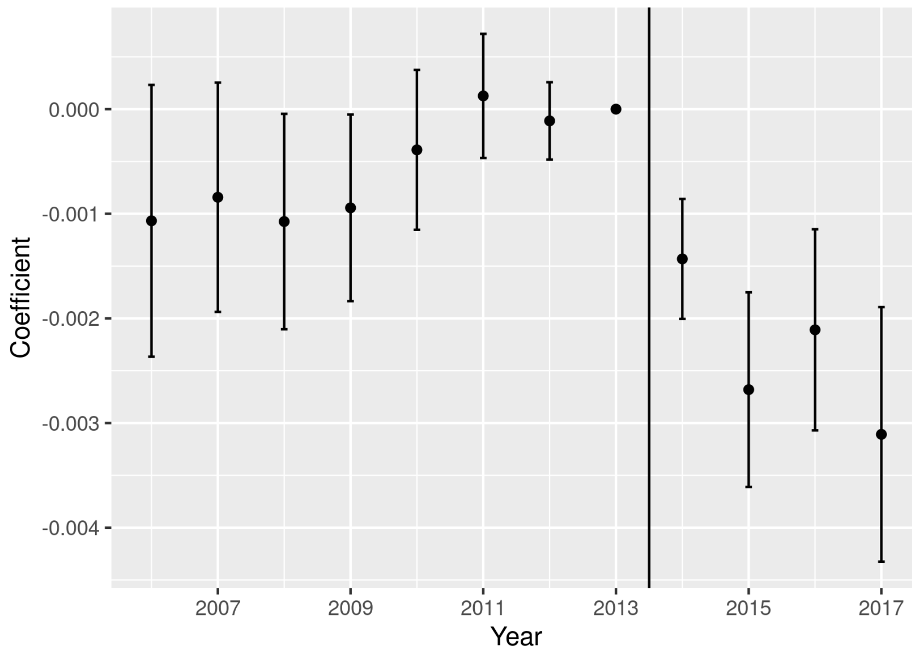
Because I have the property tax data for households moving within Philadelphia, I check their property tax before and after the move. For homeowners who moved between 2014 and 2017 within Philadelphia, their average property tax change was 75.4, 85.5% lower than the average tax increase households got. For homeowners who moved between 2015 and 2017, their property tax dropped $276.8 on average. This is more than 10% of the average property tax level in Philadelphia. These households faced the tax increase in 2014 and chose to move-out after that. The data show that they moved to properties with lower taxes.

1.4.3 Homeownership

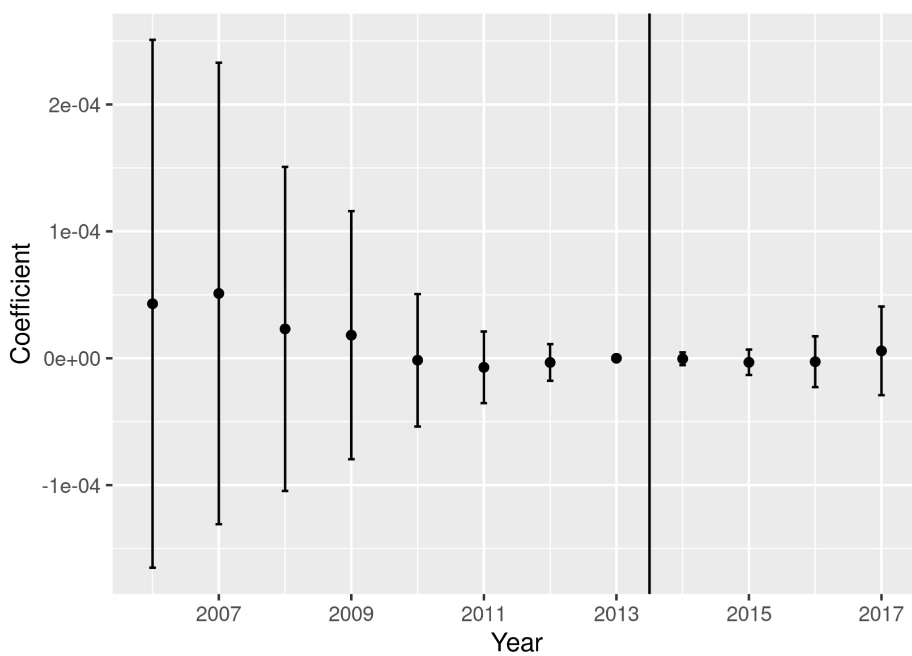
Using ownership status as the outcome variable, the event study results are shown in Figure 1.9. The figure shows no pre-trend before the reform both for owners and renters. For owners, the ownership rate decrease with property tax increases since 2014. Property tax change does not significantly affect the ownership rate of renters.

Figure 1.9: Ownership Event Study

(a) Owners



(b) Renters



Notes: Figure shows estimated coefficients (and 95% confidence intervals) on property tax increase ($1,000s) for each year from the difference-in-difference regression of homeownership on the property tax increase, property fixed effects, and year fixed effects. The 2013 coefficient is normalized to 0.

Table 1.3, shows that 1000$ increase in property tax increased the probability that an owner became a renter by 0.19 percentage point, which is 5% of all households that have ever been renters in the sample period. The impact does not sum up to zero because the sample covers households living in Philly in 2013. These households might stay in the city or move out.

Table 1.3: Household Level Dif-in-Dif Ownership Results

Dependent Variable:	Ownership Indicator		
	All	Owners	Renters
Model:	(1)	(2)	(3)
Tax Increase	5.968×10^{-5}**	0.0058	4.256×10^{-5}*
	(2.797×10^{-5})	(0.0041)	(2.28×10^{-5})
Tax Increase × I(YEAR≥2014)	4.935×10^{-7}	-0.0019***	-3.295×10^{-6}
	(1.916×10^{-5})	(0.0004)	(1.256×10^{-5})
Parcel FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	4,795,124	2,451,536	2,328,529
R^2	0.66245	0.43505	0.47257
Mean of Dependent Var	0.5133	0.9677	0.0377

Notes: Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. Data span 2007 through 2017. fixed effects. Standard errors are clustered by property. "Tax Increase" is the property tax increase (in $1000s) from 2013 to 2014 of the property that the household lives at in 2013. "I(YEAR≥2013)" is the post AVI reform indicator. Column (1) uses the full sample, which includes all the households living in Philadelphia in 2013. Column (2) includes households living in Philadelphia in 2013 and were homeowners. Column (3) includes households living in Philadelphia in 2013 and were renters.

The policy evaluations show that property tax did affect households' location and homeownership choices, but these analyses cannot show welfare influences. As is mentioned in the introduction, even if some households do not change their residential choices, their welfare could be affected because the price and rent adjust. To measure how social welfare changes in the property tax reform and compare the welfare in various equilibria, I will need to construct a model.

1.5 Model and Estimation

I build a static model that characterizes households' location and homeownership choices. In this model, every household chooses the neighborhood to live and rent or own a property in each period. I treat each time period as a separate market. Specifically, in each market, household i chooses among neighborhood n, property type j, and ownership o. The subscript t is omitted for simplicity.

By making the model static, I am ignoring the investment nature of the real estate. Specifically, I treat properties as housing services such that households pay living costs and enjoy the utility from using the property. Also, I am assuming any choice made one time period does not affect the utility in the future.

I follow previous studies of housing choice and durable goods (see, for example, Kiyotaki et al. (2011) and Sommer & Sullivan (2018)) and model household preferences over non-durable consumption, c_{inj} , and consumption of housing services, h_{inj} , as non-separable of the form

$$U_{inj} = A_n^i \frac{[(\frac{c_{inj}}{\sigma})^{\sigma}(\frac{[1-\phi^i(1-o_i)]h_{inj}}{1-\sigma})^{1-\sigma}]^{1-\rho}}{1-\rho} e^{\bar{\epsilon}_{inj}} \tag{1.3}$$

A_n^i denotes how much the local amenities in neighborhood n would affect i's utility. The parameter $\rho > 0$ is the coefficient of relative risk aversion, and $\sigma \in (0,1)$ reflects the share of consumption of goods (rather than housing services) in total expenditure. o_i is an indicator of owning a house. I assume that when household i rents a house rather than owning and controlling the same house as an owner-occupier, she enjoys a different utility by a factor ϕ^i. This disadvantage of rented housing reflects the tenant's limited discretion over the way the house is used and modified according to her tastes. It is heterogeneous across households. $e^{\bar{\epsilon}_{inj}}$ represents the idiosyncratic utility effect. $\bar{\epsilon}_{inj}$ follows Type 1 Extreme Value distribution.

Assume each household consumes 1 unit of property, but the properties are differentiated by their sizes. Type j property brings g_j housing service. So $h_{inj} = g_j$.

In neighborhood n, the rent of type j property is r_{nj} and the price of type j property is p_{nj}. Housing purchases can be financed through mortgage borrowing. Assume there is no down payment for house purchasing. Mortgage payment for type j house in neighborhood n is m_{nj}. Besides, mortgages are fully transferable, so households only need to pay the mortgage when they own the property. For simplicity, All properties are financed by a fixed term mortgage with an interest rate r^m. The mortgage payment each year m_{inj} is a fixed proportion of price bp_{nj}.

To consider transaction cost or moving cost, we will need a dynamic model. In the current static model, I assume there is no transaction cost or moving cost. The budget constraint is

$$c_{inj}q_n + (bp_{nj} + \tau_{nj})o_i + r_{nj}(1 - o_i) \leq w_i - \tau^w w_i \tag{1.4}$$

The left-hand side is the sum of expenses on nondurable goods and housing service. q_n is the price index of nondurable goods in neighborhood n. If the household owns the property, it pays the mortgage bp_{nj} and property tax τ_{nj}. Renter needs to pay rent. The right-hand side is the income net of income tax. With a flat tax rate τ^w, income tax is the tax rate multiplied with income net of deductions.

Plug the budget constraint and $h_{inj} = g_j$ back to Equation 1.3

$$U_{inj} = A_n^i \frac{[(\frac{w_i - \tau^w w_i - (bp_{nj} + \tau_{nj})o_i - r_{nj}(1-o_i)}{\sigma q_n})^{\sigma}(\frac{[1-\phi^i(1-o_i)]g_j}{1-\sigma})^{1-\sigma}]^{1-\rho}}{1-\rho} e^{\bar{\epsilon}_{inj}} \tag{1.5}$$

Taking logs and rearranging:

$$
\begin{aligned}
ln(U_{inj}) =& ln(A_n^i) \\
& + (1-\rho)\sigma ln\{w_i - \tau^w w_i - (bp_{nj} + \tau_{nj})o_i - r_{nj}(1-o_i)\} \\
& + (1-\rho)(1-\sigma)ln[1-\phi^i(1-o_i)] + (1-\rho)(1-\sigma)lng_j - (1-\rho)\sigma ln(\sigma q_n) \\
& - (1-\rho)(1-\sigma)ln(1-\sigma) - ln(1-\rho) + \bar{\epsilon}_{inj}
\end{aligned} \tag{1.6}
$$

Let

$$
A_n^i = \prod_a (X_n^a)^{\bar{\alpha_a^i}} \tag{1.7}
$$

where X_n^a is type a amenity in neighborhood n.

$$
\begin{aligned}
(1-\rho)\sigma =& \bar{\mu} \\
\bar{\beta_i} =& (1-\rho)(1-\sigma)ln(1-\phi^i) \\
\bar{\xi_n} =& -(1-\rho)\sigma ln(\sigma q_n) \\
\bar{\eta_j} =& (1-\rho)(1-\sigma)lng_j \\
\bar{d} =& -(1-\rho)(1-\sigma)ln(1-\sigma) - ln(1-\rho)
\end{aligned} \tag{1.8}
$$

. The right-hand side of Equation 1.6 becomes

$$
\begin{aligned}
& \bar{d} + \bar{\mu} ln\{w_i - \tau^w w_i - (bp_{nj} + \tau_{nj})o_i - r_{nj}(1-o_i)\} \\
& + \sum_a \bar{\alpha_a^i} ln(X_n^a) + \bar{\beta_i}(1-o_i) + \bar{\xi_n} + \bar{\eta_j} + \bar{\epsilon}_{inj}
\end{aligned} \tag{1.9}
$$

Finally, I will divide the previous equation by the variance of the shock $\bar{\epsilon}_{inj}$ to normalize it to 1. The error ϵ_{inj} follows iid. Logistic (0, 1) distribution. After such normalization, the expression for the indirect utility is

$$
\begin{aligned}
u_{inj} + \epsilon_{inj} =& d + \mu ln\{w_i - \tau^w w_i - (bp_{nj} + \tau_{nj})o_i - r_{nj}(1-o_i)\} \\
& + \sum_a \alpha_a^i ln(X_n^a) + \beta_i(1-o_i) + \xi_n + \eta_j + \epsilon_{inj} \\
\alpha_a^i =& \alpha_a^1 o_i + \sum_k \alpha_{ak}^2 Z_i^k + \alpha_a^3 \nu_i \\
\beta_i =& \beta^1 + \sum_k \beta_k^2 Z_i^k + \beta^3 \omega_i
\end{aligned} \tag{1.10}
$$

Z_i^k is the kth type dummy of household i. Assume the coefficient α_a^i is a linear function of household characteristics, with different coefficients for renters and owners. It characterizes households' heterogeneous preference on neighborhood amenities. β_i characterizes households' heterogeneous preference on renting. α_a^i and β_i also contain ν_i and ω_i that characterize the unobserved heterogeneous preferences. Assume ν_i and ω_i follow standard normal distribution. ξ_n is the unobservable neighborhood quality.

Use subscript o to denote the ownership choice ($o = 1$ or 0). p_{njo} is the price or rent of type j house in neighborhood n with ownership choice as o. Then the utility that household i chooses n, j, and o is

$$
\begin{aligned}
u_{injo} + \epsilon_{injo} =& d + \mu ln\{w_i - \tau^w w_i - (bp_{njo} + \tau_{njo})o - p_{njo}(1-o)\} \\
& + \sum_a \alpha_a^1 ln(X_n^a) + \sum_a \sum_k \alpha_{ak}^2 Z_i^k ln(X_n^a) + \sum_a \alpha_a^3 \nu_i ln(X_n^a) \\
& + \beta^1(1-o) + \sum_k \beta_k^2 Z_i^k (1-o) + \beta^3 \omega_i (1-o) \\
& + \xi_{njo} + \epsilon_{injo}
\end{aligned} \tag{1.11}
$$

Use δ_{njo} to denote the homogeneous part in the utility function and λ_{injo} for the rest. Let

$$
\delta_{njo} = \sum_a \alpha_a^1 ln(X_n^a) + \beta^1(1-o) + \xi_{njo} \tag{1.12}
$$

$$
\begin{aligned}
\lambda_{injo} =& \mu ln\{w_i - \tau^w w_i - (bp_{njo} + \tau_{njo})o - p_{njo}(1-o)\} \\
&+ \sum_a \sum_k \alpha_{ak}^2 Z_i^k ln(X_n^a) + \sum_a \alpha_a^3 \nu_i ln(X_n^a) \\
&+ \sum_k \beta_k^2 Z_i^k (1-o) + \beta^3 \omega_i (1-o)
\end{aligned}
\tag{1.13}
$$

If the financial cost of living in a house exceeds the household's after-tax income, the household has zero probability to choose the house. Use F_i to denote the set of n and j that household i could afford. The probability that household i chooses n, j, o would be

$$
P_{injo} = \frac{exp\{\delta_{njo} + \lambda_{injo}\}}{\sum_{n',j',o' \in F_i}(exp\{\delta_{n'j'o'} + \lambda_{in'j'o'}\})} \tag{1.14}
$$

Market clearing conditions are

$$
\sum_i P_{injo} = S_{njo} \quad \forall n, \forall j, \forall o \tag{1.15}
$$

I assume the market supply S_{njo} is exogenous. The Licenses and Inspections Building and Zoning Permits dataset from the Philadelphia government record the construction plans and building inspections reviewed by the Department of Licenses Inspections in Philadelphia. Table 1.4 shows the number of new constructions and the growth rate from 2010 to 2015. The reform in 2014 does not change the trend of the new construction growth rate. Therefore, the fixed supply assumption aligns with the data fact.

Table 1.4: New Constructions in Philadelphia

Year	New Construction	Growth Rate
2010	938	
2011	805	-14.18%
2012	1039	29.07%
2013	1180	13.57%
2014	1294	9.66%
2015	1352	4.48%

Notes: Data comes from The Licenses and Inspections Building and Zoning Permits dataset from Philadelphia government. Table shows the number of new constructions and the growth rate from 2010 to 2015.

1.5.1 Estimation Strategy

For simplicity, I start with estimating a model without product types, or unobserved heterogeneity. There are two household types: high-wealth and low-wealth. I group households by income into three bins and use the maximum income in each bin as the group income. [1]

Under the simplification assumptions, the indirect utility function becomes

$$\begin{aligned} u_{ino} + \epsilon_{ino} =& d + \mu ln[(1-\tau^{w})w_i] - (bp_{no} + \tau_{no})o - p_{no}(1-o)] \\ &+ \beta^1(1-o) + \sum_k \beta^2 Z_i(1-o) + \xi_{no} + \epsilon_{ino} \end{aligned} \tag{1.16}$$

I estimate parameters using simulated methods of moments. The moment conditions include a set of aggregate moments that match the market demand shares, a set of micro

1. The utility function has log consumption. To make sure all households have non-negative consumption, I need to take group income $\geq$ the highest income in the group.

moments to approximate the heterogeneity of preferences, and a set of IV moments to solve for the endogeneity of price.

The aggregate moment condition is

$$G^1(\theta) = S_{no} - \sum_i P^i_{no}(\theta) \quad \forall n, o \tag{1.17}$$

Micro moments include the expectation of price for each income group and the expectation of ownership for each income and household type group.

$$G^2_{pw}(\theta) = \sum_{i \in w_g} p_i - \sum_{i \in w_g} \sum_{no} P^i_{no}(\theta) p_{no} \quad \forall w_g \tag{1.18}$$

$$G^2_{ow}(\theta) = \sum_{i \in w_g \cap \{Z^k_i = 1\}} o_i - \sum_{i \in w_g \cap \{Z^k_i = 1\}} \sum_{no} P^i_{no}(\theta) o \quad \forall w_g, k \tag{1.19}$$

The instrument variable moment is

$$G^3(\theta) = IV' \cdot \xi \tag{1.20}$$

I follow Neilson (2020) and apply the modified MPEC approach with micro moments and normalization restrictions of ξ.

1.5.2 Instrumental Variable

AVI was adopted in 2013 and took effect in 2014 in Philadelphia. Most property prices were rising from 2000 to 2010s. Before AVI (in 2014), the longer a property had not been assessed, the lower its assessment value would be compared with its real market value. Those properties, with lower property tax, were more likely to have higher prices.

For all properties that have been transacted before 2014, I found when they were last assessed and created a variable "Assessment Lag" A_{jt}. Denote the price of property j in

Year t with P_{jt}. Table 1.5 shows the results of the following regression:

$$P_{jt} = \alpha_0 + \alpha_1 A_{jt} + \gamma_t + \epsilon_{jt} \tag{1.21}$$

where γ_t is the year fixed effect.

Table 1.5: Unit Level First Stage Using Pre 2014 Sample

Dependent Variable:	Price
Model:	(1)
Variables	
Assessment Lag	7,132.7***
	(681.8)
Fixed-effects	
Year	Yes
Fit statistics	
Standard-Errors	Year
Observations	62,182
R^2	0.00534
Within R^2	0.00438

Notes: Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. The regression uses Infogroup property dataset and covers all properties that have been transacted before 2014. Table shows the regression results of price on the time difference between the transaction year and the year when that property was last assessed and year fixed effect.

The significantly positive coefficient in front of the assessment lag variable shows strong correlations in the first stage.

To satisfy the exclusion restriction, the instrumental variable needs to be uncorrelated with the unobserved neighborhood quality. In some jurisdictions, properties were reassessed in a transaction or after construction and renovation. Combining the transaction dataset and assessment dataset, I find a very low correlation between transaction and reassessment (0.01). Besides, I collect zoning and building permit information from 2007 to 2015 from the Philadelphia government and find the correlation between a completed permit and reassessment is 0.017.

These tests show that reassessment occurrence was not highly correlated with transaction or construction. In addition, I find assessment lag does not significantly correlate with neighborhood median income, median age, post-secondary education rate, and white people proportion.

The assessment process in Philadelphia, in fact, was very obscure. The Philadelphia Citizen, a non-profit, non-partisan media organization, says that the assessment process of Philadelphia is "mysterious".[2] Households could hardly get information and predict when a property would be assessed. So, the assessment lag would not correlate with factors observable to households and unobservable to econometricians that affected households' residential choices. It is reasonable to assume that when the property was assessed before 2014 was exogenous.

In the estimation, I aggregate the data up to the Zip-code level. Use L_{nt} to denote the average year difference between the previous assessment and the current year t for properties in neighborhood n. The IV is defined as the interaction of L_{nt} and o:

$$Z_{not} = L_{nt} \times o \tag{1.22}$$

It passes the first stage test on both Zip-code level and Census tract level (Table B.1). In addition, to deal with the concern that newer neighborhoods could have more constructions

2. "Your city defined - the office of property assessment" https://thephiladelphiacitizen.org/citizencast-your-city-defined-the-office-of-property-assessment/

and shorter average assessment lags, I control the average built year of properties in each neighborhood. The assumption of the IV is that conditional on the average age of the neighborhood and ownership status, when the properties were last assessed was uncorrelated with the current year's unobserved quality.

1.5.3 Identification

In the static choice model, I identify the parameters using the household sorting patterns. The proportion of households in each neighborhood owning properties helps identify δ_{ino}. If a neighborhood has a low ownership rate, the estimation will result in a low δ.

The moments derived from the household level data help identify μ, β^1, and β^2. The average ownership for each household type group helps identify the heterogeneous utility from owning. Specifically, in the baseline estimation where household type is characterized by an indicator of wealth level, the ownership rate for the low-wealth group is informative about β^1 and that for the high-income group is informative about $\beta^1 + \beta^2$. The higher ownership rate low-wealth households have relative to the neighborhood average, the lower β^1 will be estimated. The difference in the ownership rate between high-wealth households and low-wealth households helps identify β^2. The large the homeownership rate for high-wealth households is compared with low-wealth households, the smaller β^2 will be estimated.

For coefficient μ, the expectation of ownership and property price for each income group of households are both informative. Household income, property price/rent, and ownership status together determine the consumption level. The higher the expectation of consumption level is, the larger μ I will get in the estimation. For high-income households, the extent to which their property price and ownership rate differ from lower-income households helps identify μ. If high-income households pay a lot more on housing than low-income households, it shows that households put little value on consumption, so the estimated μ will be smaller.

Finally, the set of instrumental variable moments allows me to identify ξ. These unobserved product (neighborhood and ownership combination) qualities are estimated like fixed effects.

The ratio of annual housing payment to property price b is not identifiable in this model. Therefore, I assume all households have a 30-year term mortgage with a fixed interest rate at 4% and they pay 0% down payment. Under this assumption, $b = 0.058$. I check the robustness of the results to this assumption in the next section.

Here are some other assumptions I make to simplify the computation. The neighborhood is defined by zip-code area. Price, rent, and property tax are all zip code-year averages. Assume there is no income tax so that $\tau_w = 0$.

1.6 Results and Analysis

1.6.1 Baseline Estimation Results

I use the sample of households that lived in neighborhoods with price/rent information in 2012 to do the estimation. There are 623,455 households in the sample choosing among 91 products (combination of neighborhood and ownership). Products that no households had ever chosen in the sample were dropped.

I define neighborhoods by Zip-code area. The neighborhood-year price is the average price of houses transacted in that year. The neighborhood-year rent is the average rent of multifamily units in that year. Assume $b = 0.058$.

Households are grouped into two types by wealth level.

Under these assumptions, the utility function is simplified to

$$\begin{aligned} u_{ino} + \epsilon_{ino} =& d + \mu ln[(1-\tau^{w})w_i] - (bp_{no} + \tau_{no})o - p_{no}(1-o)] \\ &+ \beta^1(1-o) + \beta^2 Z_i(1-o) + \xi_{no} + \epsilon_{ino} \end{aligned} \tag{1.23}$$

where Z_i is an indecator of high-wealth households. The parameters to be estimated are μ,

β_1, β_2.

The estimates are shown in Table 1.6.

Table 1.6: Baseline Estimates Using 2012 Sample

Parameter		Coeff.	Std Err.
μ	utility from log consumption	6.5145	(3.1995)
β^1	utility from renting (low-wealth)	1.0207	(0.1659)
β^2	utility difference from renting (high vs. low-wealth)	-2.6917	(0.9014)

Notes: Table shows estimates of the baseline model. Parameters are estimated with method of moments.

The result indicates that households prefer consumption. High-wealth households prefer owning and low-wealth households prefer renting. The magnitude of utility on renting of low-wealth households and the disutility on renting of high-wealth households are very close. The utility of log consumption is much larger. For a high-wealth household with income \$100,000 and \$10,000 annual housing payment, conditional on housing payment, the extra utility of owning a property compared with renting provides the same utility gain as having \$39,000 money.

Table 1.7 shows the true and predicted moments, which shows the model fit.

Table 1.7: Model Fit

Moment	Data	Model Predicted
Average Price/Rent of Low Income	29,650	34,350
Average Price/Rent of Middle Income	48,630	53,360
Average Price/Rent of High Income	131,270	121,600
Ownership Rate of Low Income, Low Wealth	22.21%	22.54%
Ownership Rate of Low Income, High Wealth	72.53%	76.69%
Ownership Rate of Middle Income, Low Wealth	15.70%	16.98%
Ownership Rate of Middle Income, High Wealth	69.20%	69.81%
Ownership Rate of High Income, Low Wealth	16.26%	15.60%
Ownership Rate of High Income, High Wealth	70.01%	67.64%

Notes: Table shows the model fit of the baseline model. It compares the data moments and model predicted moments.

1.6.2 Robustness Checks

Neighborhood Characteristic

In the baseline estimation, I ignore neighborhood characteristics. Households might have heterogeneous preferences on neighborhood characteristics. If we consider that in the estimation, will the estimates be affected? To answer this question, I get the median income data for each zip-code area in Philadelphia in 2012 from American Community Survey 5-Year Data and use that as the neighborhood characteristic.

In this estimation, the utility function is

$$\begin{aligned} u_{ino} + \epsilon_{ino} =& d + \mu ln\{w_i - \tau^w w_i - (bp_{no} + \tau_{no})o - p_{no}(1-o)\} \\ & + \alpha^1 ln(X_n) + \alpha^2 Z_i ln(X_n) \\ & + \beta^1 (1-o) + \beta^2 Z_i (1-o) \\ & + \xi_{no} + \epsilon_{ino} \end{aligned} \tag{1.24}$$

where $ln(X_n)$ is the median income of each neighborhood (neighborhood characteristic) and Z_i is an indicator of high-wealth household (household characteristic). The neighborhood median income is in the unit of \$100,000. The parameters to be estimated are μ, α^1, α^2, β^1, β^2. To identify α^1 and α^2, I add one more set of moment conditions: the expectation of neighborhood characteristic for each household type.

$$G_x^2(\theta) = \sum_{i \in \{Z_i^k = 1\}} X_i - \sum_{i \in \{Z_i^k = 1\}} \sum_{no} P_{no}^i(\theta) X_n \quad \forall k \tag{1.25}$$

Table 1.8 shows the estimation results.

Table 1.8: Robustness Check: Estimation with Neighborhood Characteristic

Parameter		Coeff.	Std Err.
μ	utility from log consumption	6.7751	(5.8791)
β^1	utility from renting (low-wealth)	1.2111	(0.3072)
β^2	utility difference from renting (high vs. low-wealth)	-2.6926	(0.2102)
α^1	utility from neighborhood income (low-wealth)	0.0064	(7.2050)
α^2	utility difference from neighborhood income (high vs. low-wealth)	-1.4536	(5.5678)

Notes: Table shows the estimates of the model with heterogeneous utility on the neighborhood characteristic (income). Parameters are estimated with method of moments.

The estimates of preference on the neighborhood characteristic are not significant. Adding these to the model does not affect the estimates of the main parameters much.

Unobserved Heterogeneity

In the baseline estimation, I assume households have homogeneous preference on homeownership. To check whether adding heterogeneous preference will affect the estimation results, I estimate a model with heterogeneous preference on homeownership. The utility function is

$$\begin{aligned} u_{ino} + \epsilon_{ino} =& d + \mu ln[(1-\tau^w)w_i] - (bp_{no} + \tau_{no})o - p_{no}(1-o)] \\ &+ \beta^1(1-o) + \beta^2 Z_i(1-o) + \beta^3\omega_i(1-o) + \xi_{no} + \epsilon_{ino} \end{aligned} \tag{1.26}$$

The probability that household i chooses n, o is

$$P_{ino} = \int \frac{exp\{\delta_{no} + \lambda_{ino} + \beta^3\omega_i(1-o)\}}{\sum_{n',o' \in F_i}(exp\{\delta_{n'o'} + \lambda_{in'o'} + \beta^3\omega_i(1-o)\})}\, dF_\omega(\omega_i) \tag{1.27}$$

Without loss of generality, the distribution of the unobservable characteristic is assumed to be normal with a zero mean and a standard deviation of β^3 so that ω_i follows a standard normal distribution. For each household, I simulate 100 values of ω_i for the calculation of integral. For simplification, I estimate the model with a random subsample of 10,000 households.

Table 1.9 shows the estimation results. The coefficient on unobserved heterogeneous preference on homeownership is insignificant. Adding the unobserved heterogeneity also affects the significance of other coefficients. But the sign of the coefficients is consistent with those in the baseline estimates.

Table 1.9: Robustness Check: Estimation with Unobserved Heterogeneity

Parameter		Coeff.	Std Err.
μ	utility from log consumption	12.4046	(6.4247)
β^1	utility from renting (low-wealth)	6.6924	(14.5293)
β^2	utility difference from renting (high vs. low-wealth)	-10.4309	(50.6070)
β^3	utility from unobserved heterogeneity	6.9560	(36.0432)

Notes: Table shows the estimates of the model with unobserved heterogeneity. Parameters are estimated with method of simulated moments.

Sensitivity to b

In the baseline estimation, I assume $b = 0.058$, which is calculated by the rate of mortgage payment to price under a 30-year fixed interest rate mortgage with rate 4% and zero down payment. In this section, I check how sensitive my estimates are to the value of b.

I collect mortgage data from Freddie Mac. The dataset includes the interest rate, income, and loan value for all the mortgages in Philadelphia in 2012. Because most people get an interest rate lower than 4%, their b values are lower than 0.058. I calibrate the b by household income and get the average b for my sample is 0.054. Assume $b = 0.054$, I get the estimates in 1.10. The results are very similar to the baseline results.

Table 1.10: Robustness Check: $b = 0.054$

Parameter		Coeff.	Std Err.
μ	utility from log consumption	6.0304	(3.6013)
β^1	utility from renting (low-wealth)	1.0635	(0.2081)
β^2	utility difference from renting (high vs. low-wealth)	-2.7772	(0.9562)

Notes: Table shows the estimates of the model with b value being 0.054. Parameters are estimated with method of moments.

I also calculate the b for different income groups. Because higher-income households usually have higher credit scores and can get lower lending interest rates. The higher the income is, the lower the b is. The b for income groups from the lowest to the highest are 0.0543, 0.0542, 0.0535. Using these values in the estimation, I get the results in 1.11. Still, the results are very similar to the baseline.

Table 1.11: Robustness Check: Different bs by Income Group

Parameter		Coeff.	Std Err.
μ	utility from log consumption	6.0752	(3.5894)
β^1	utility from renting (low-wealth)	1.0632	(0.2050)
β^2	utility difference from renting (high vs. low-wealth)	-2.7702	(0.9590)

Notes: Table shows the estimates of the model with different b value for different income groups. The b for income groups from the lowest to the highest are: 0.0543, 0.0542, 0.0535. Parameters are estimated with method of moments.

To conclude, the estimates are robust to models with or without neighborhood characteristics, unobserved heterogeneous preference on homeownership, and different values of b.

1.6.3 Counterfactual Analysis

I use the baseline model to analyze a counterfactual case where all properties have a flat tax rate 1.34%. That tax rate is the rate in Philadelphia's policy. Ideally, the real property tax should be the true property value times the tax rate, and all properties are taxed at the same rate. My goal is to measure the social welfare in that ideal case and compare it with that under the unequal property tax system in reality.

In the current property tax system, the effective tax rate is regressive. Higher value properties are taxed at a lower rate compared with lower value properties. When adjusting the system to a flat-rate one, we would expect higher value properties to get a tax increase and lower value properties to get a tax deduction. The property tax change will drive the price of higher value properties up and lower value properties down. Households will relocate and change their homeownership decision accordingly.

I first find equilibrium prices under the counterfactual system. Then I simulate households' choices and compute the ownership rate for each type of household, shown in Table 1.12

Table 1.12: Counterfactual Ownership Rate by Household type

Type	Original	Counterfactual
Low income, low wealth	0.2302	0.2350
Low income, high wealth	0.7717	0.7764
Middle income, low wealth	0.1698	0.1709
Middle income, high wealth	0.6981	0.6997
High income, low wealth	0.1560	0.1562
High income, high wealth	0.6764	0.6767

Notes: Table shows the original and counterfactual homeownership rates for each group of households. In the original case, real property tax rate is unequal across properties. In the counterfactual case, every property is taxed at a flat rate 1.34%.

The table shows that when changing the unequal property tax system to a flat-rate system, all households are more likely to become homeowners, but the homeownership rate change for low and middle-income households is much greater than that for high-income households. This is because low and middle-income households are more likely to live in lower-value properties. A flat-rate property tax system is more likely to reduce the tax and price of those properties. So, low and middle-income households are more likely to become homeowners.

The original ownership rate for high-wealth households is higher than that for low-wealth households. But the ownership rate increase is more substantial for low-wealth households compared with high-wealth households. Even though low-wealth households prefer renting to owning, the property tax decrease and price decrease make more low-wealth households better off owning than renting, which means that the increase of utility from consumption exceeds the utility loss from homeownership.

How welfare changes for different types of households, however, remains a question. Low-income and middle-income households are more likely to live in low-value properties. Among these households, because high-wealth ones prefer owning, they are more likely to own those low-value properties. When the property tax of low-value houses decreases, more households want to become owners. Because high-wealth households prefer owning, some of them may enjoy welfare increases caused by homeownership. However, housing prices will increase as a response to the increasing demand. These households might be hit by the decreasing consumption as a result. Which impact is larger determines which direction the welfare will go. At the same time, rent drops because of the decreasing demand. Low-wealth, low-income, and middle-income households are more likely to be renters of these low-value properties. The fact that more of these households become owners shows that the rent reduction is not as much as the net reduction of tax and price. But these households may still gain welfare because of the corresponding consumption increase.

For high-income households who are more likely to live in high-value properties, how the welfare change is also uncertain. When property tax increases, some households might not be able to afford houses and become renters. For high-wealth households, because they prefer owning, they will get a welfare reduction. At the same time, price of houses will adjust, which compensates for the level of consumption change brought by the tax increase. The magnitude of these two changes determine how the welfare changes for high-value property owners (likely high-income high-wealth households). At the same time, rent increases, could hurt high-income low-wealth households.

In addition to the distributional impact measured by the welfare change for each group, we are also interested in knowing how the overall welfare changes across the whole population and the tax amount change.

I calculate the welfare change for different types of households. Welfare is defined by the aggregation of household utility and property tax. Table 1.13 shows the welfare change when we modify the property tax system from a regressive one to an equal one.

Table 1.13: Welfare Change from the Real Property Tax to Equal Rate Property Tax

Type	Welfare Change
Low income, low wealth	4.59%
Low income, high wealth	3.70%
Middle income, low wealth	9.23%
Middle income, high wealth	4.28%
High income, low wealth	-0.27%
High income, high wealth	-0.17%

Notes: Table shows the welfare change for each type of households if the property tax system is changed from the current one to the flat rate one (1.34%).

The results show that when we change the property tax system to an equal rate one, the welfare of low and middle-income households increases. The welfare of high-income households decreases, but not as much as the increase from low and middle-income households. Among households at the same income level, the welfare change of high-wealth households is smaller than that of low-wealth households.

For high-wealth, low-income, and middle-income households that are more likely to own low-value properties, the increase of welfare shows that the gain from more ownership exceeds the loss of price increase. The welfare gain of low-wealth, low-income, and middle-income households is more likely to come from the consumption gain caused by rent decrease.

For high-income high-wealth households, the welfare loss shows that the loss from ownership exceeds the gain from the price drop. Low-wealth households are hit by the rent increase.

Overall, the aggregate household welfare increase is 3.53%. Besides, the total property tax increases by 23.64%.

The result shows that when the regressivity of the property tax system is corrected, not only the government will collect a higher property tax income, but also the welfare of the whole population can be improved. Only high-income households will be hurt by the change, but the welfare reduction is at a very small magnitude.

1.7 Conclusion

In this paper, I measure the impact of the current regressivity in the property tax system and find that correcting the regressivity and transforming the property tax system to a flat-rate one will improve households' welfare by 3.53%.

I start with showing how property tax affected households' homeownership and location decisions. Using a natural experiment, the AVI reform in Philadelphia, I find that 1000$ more property tax increased the probability that owners moved out by 0.28 percentage points, which was 15% increase relative to the average migration rate. It also increased the renter's migration probability, but with a much smaller magnitude and significance. Besides, 1000$ increase in property tax increased the probability that an owner became a renter by 0.19 percentage point, which is 5% of all households that have ever been renters in the sample period.

Finding that property tax can both affect households' location choice and homeownership choice, I built a demand model which characterizes these joint decisions. Under some simplification assumptions, I estimate the model with Philadelphia housing market data and household migration history data with simulated methods of moments approach. I also validated the robustness of the estimates with neighborhood characteristics, unobserved heterogeneity, and different assumptions on the relationship between annual payment and property price.

In the end, I analyze a counterfactual scenario where all properties are taxed at a flat rate. I find that low-income and middle-income households will enjoy a welfare increase while high-income households will be slightly hurt by the rising tax. The overall population will gain a 3.53% welfare increase. Meanwhile, the government will collect 23.64% more property tax through this reform.

This study indicates that the existing regressivity hurts the households and the government. Besides, it creates strong inequity on social welfare. Correcting the regressivity and creating a flat-rate property tax system will be beneficial to the whole society.

REFERENCES

Almagro, M. & Domínguez-Iino, T. (2021), 'Location sorting and endogenous amenities: Evidence from amsterdam'.

Bayer, P., McMillan, R., Murphy, A. & Timmins, C. (2016), 'A dynamic model of demand for houses and neighborhoods', *Econometrica* **84**, 893–942.

Berry, C. (2021), 'Reassessing the property tax', *Discussion Papers* .

Binner, A. & Day, B. (2015), 'Exploring mortgage interest deduction reforms: An equilibrium sorting model with endogenous tenure choice', *Journal of Public Economics* **122**, 40–54.

Diamond, R., McQuade, T. & Qian, F. (2019), 'The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco', *American Economic Review* **109**(9), 3365–94.

Ding, L. & Hwang, J. (2020), 'Effects of gentrification on homeowners: Evidence from a natural experiment', *Regional Science and Urban Economics* **83**.

Kiyotaki, N., Michaelides, A. & Nikolov, K. (2011), 'Winners and losers in housing markets', *Journal of Money, Credit and Banking* **43**(2/3), 255–296.

Neilson, C. A. (2020), 'Targeted vouchers, competition among schools, and the academic achievement of poor students'.

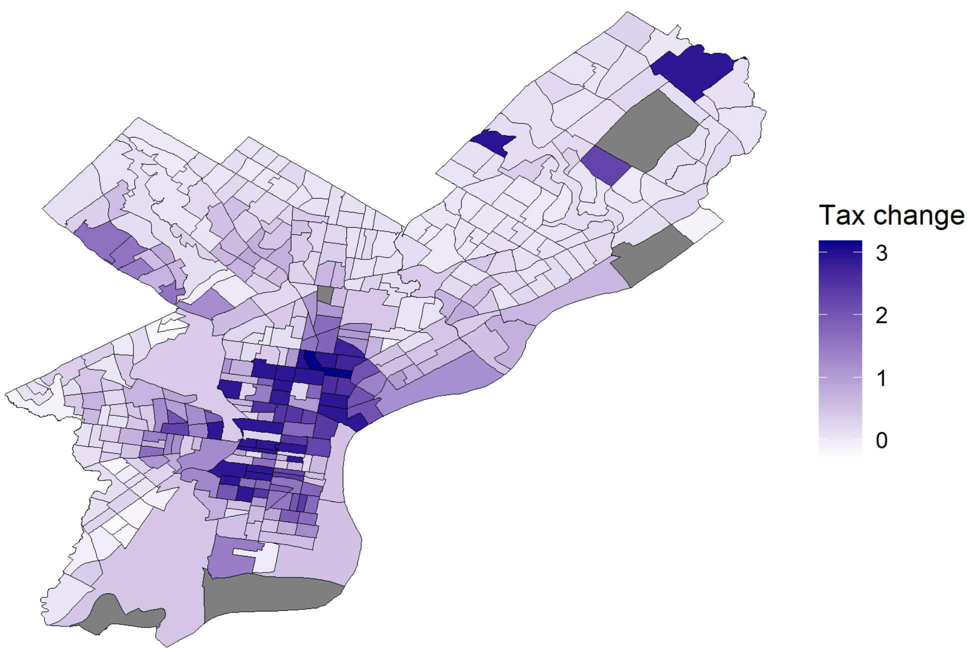
Shan, H. (2010), 'Property taxes and elderly mobility', *Journal of Urban Economics* **67**(2), 194–205.

Sommer, K. & Sullivan, P. (2018), 'Implications of us tax policy for house prices, rents, and homeownership', *American Economic Review* **108**(2), 241–274.

APPENDIX A

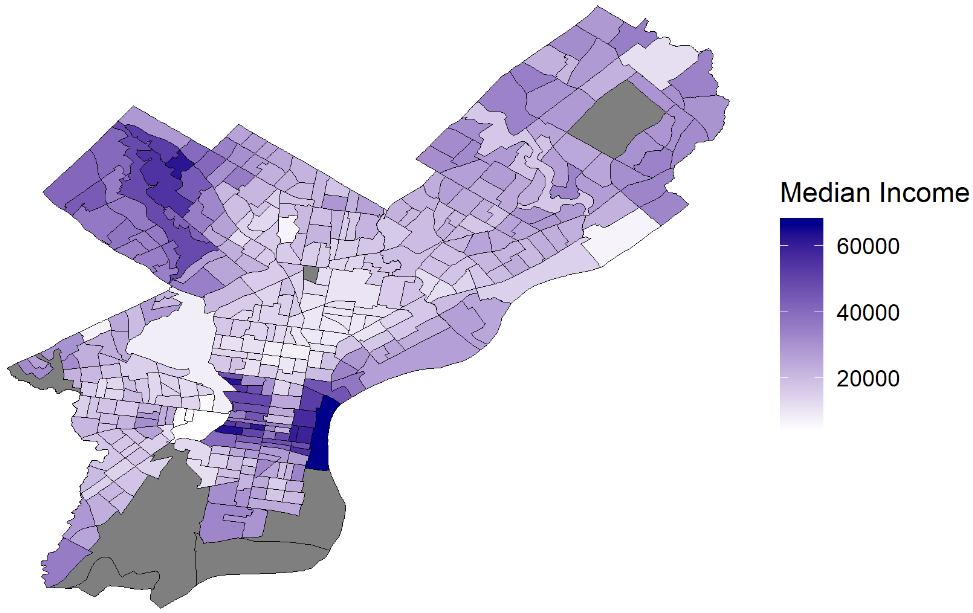
ADDITIONAL FIGURES

Figure A.1: Geographical Variation of Average Tax Increase Rate



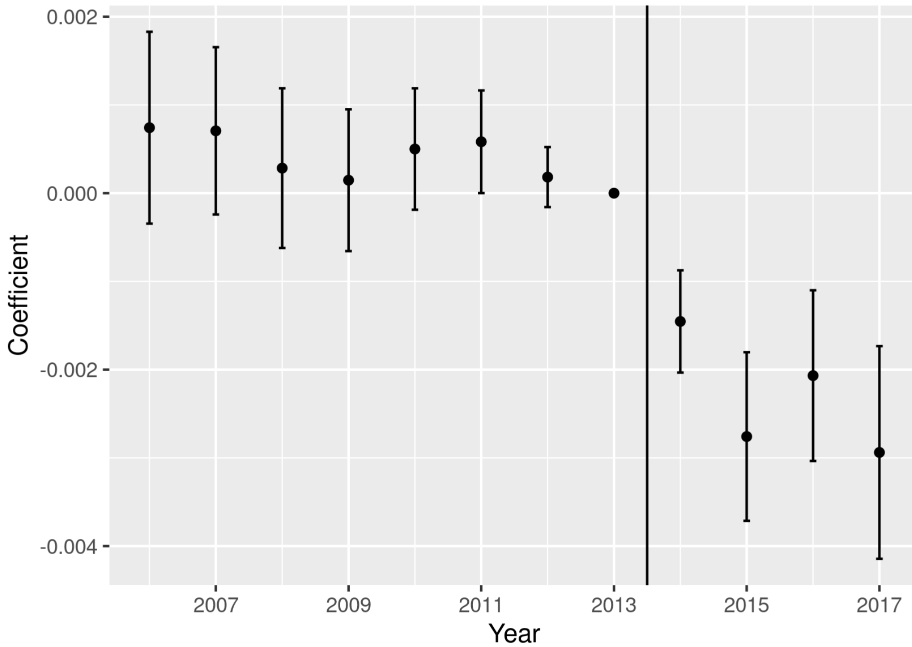
Notes: Figure shows the average property tax increase from 2013 to 2014 for each Census tract in Philadelphia. Data come from American Community Survey.

Figure A.2: Average Median Income by Census Tract (2010-2017)



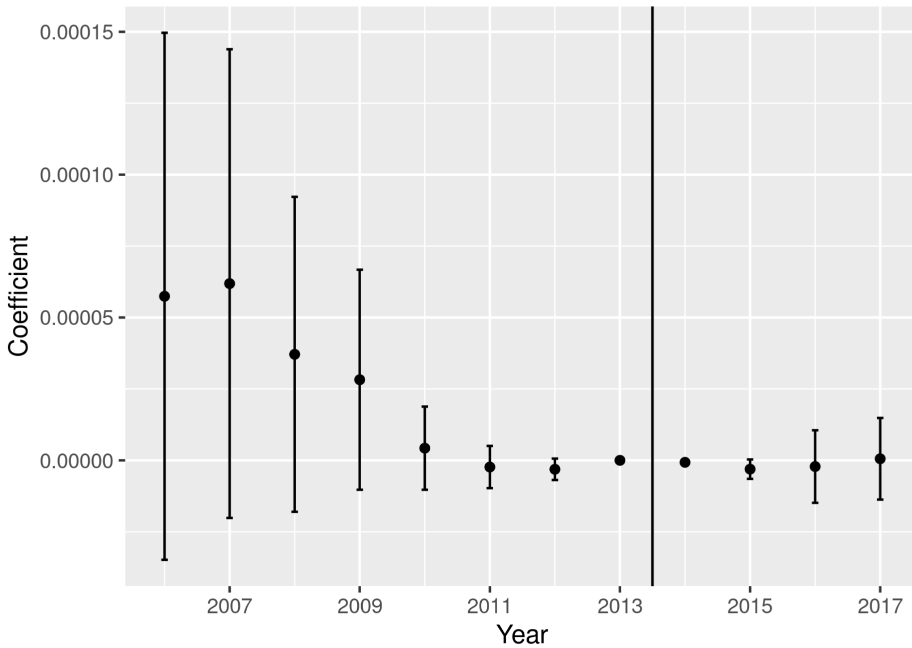
Notes: Figure shows the median annual income in 2013 for each Census tract in Philadelphia. Data come from American Community Survey.

Figure A.3: Household Level Ownership (Owners in 2013, 3-way FE)



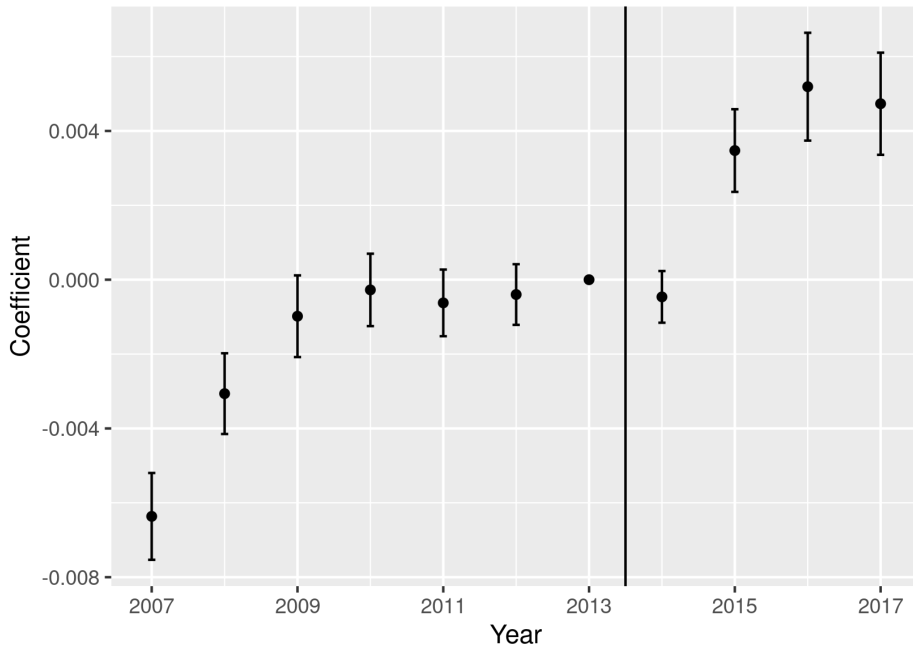
Notes: Figure shows estimated coefficients (and 95% confidence intervals) on property tax increase ($1,000s) for each year from the difference-in-difference regression of homeownership on property tax increase, property fixed effects, household fixed effects, and year fixed effects. Sample includes all households that lived in Philadelphia in 2013 and were owners. The 2013 coefficient is normalized to 0.

Figure A.4: Household Level Ownership (Renters in 2013, 3-way FE)



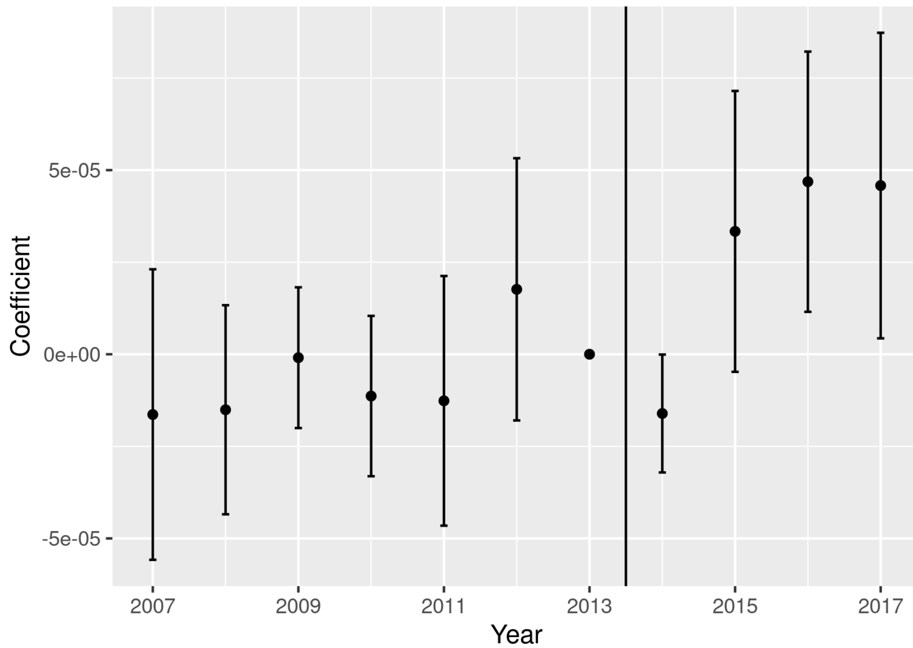
Notes: Figure shows estimated coefficients (and 95% confidence intervals) on property tax increase ($1,000s) for each year from the difference-in-difference regression of homeownership on property tax increase, property fixed effects, household fixed effects, and year fixed effects. Sample includes all households that lived in Philadelphia in 2013 and were renters. The 2013 coefficient is normalized to 0.

Figure A.5: Household Level Move-out (Owners in 2013, 3-way FE)



Notes: Figure shows estimated coefficients (and 95% confidence intervals) on property tax increase ($1,000s) for each year from the difference-in-difference regression of moving indicator on property tax increase, property fixed effects, household fixed effects, and year fixed effects. Sample includes all households that lived in Philadelphia in 2013 and were owners. The 2013 coefficient is normalized to 0.

Figure A.6: Household Level Move-out (Renters in 2013, 3-way FE)



Notes: Figure shows estimated coefficients (and 95% confidence intervals) on property tax increase ($1,000s) for each year from the difference-in-difference regression of moving indicator on property tax increase, property fixed effects, household fixed effects, and year fixed effects. Sample includes all households that lived in Philadelphia in 2013 and were renters. The 2013 coefficient is normalized to 0.

APPENDIX B

ADDITIONAL TABLES

Table B.1: First Stage of the Estimation Sample (2012 Data)

Dependent Variable:	Price	
Model:	Zip code Level	Tract Level
Assessment Year Lag×Ownership	104,329.600***	28,757.580***
	(34,440.780)	(9,086.199)
Ownership	−487,492.500**	7,756.845
	(230,735.700)	(62,742.040)
Constant	1,016.255	1,007.252
	(23,450.930)	(12,718.350)
F-statistic of the Excluded Instruments	9.176	10.017
Observations	91	644
Adjusted R^2	0.336	0.189

Notes: Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. The regressions use Infogroup property dataset. Table shows the regression results of zip code and Census tract level average property price on homeownership rate and the interaction of ownership rate with the average time difference between 2013 and the year when that property was last assessed.

APPENDIX C

DATA

C.1 Infogroup Data

C.1.1 Data Structure

The Infogroup dataset spans from 2006 to 2017. Each observation is a household in a year. The observation shows the address and some household characteristics. For each household, the dataset records up to three household members and their names.

Infogroup Consumer Database is compiled from five types of primary sources: Utility New Connects; Real Estate Data (Tax Assessor and Deed information); Voter Registration Lists where available for marketing applications; Credit Card Transaction data; and Public Records such as hunting licenses and boat registrations.

Definition of household: A Family is defined as one or more individuals that can appear to function as an economic unit based upon surname or other logical evidence.

- If an entire family moves together, the Family ID will remain consistent.
- If part of the family moves, the Family ID will remain with the individual(s) that remains at the existing address.
- If a family divides and both sides leave the current residence the existing Family ID would stay with the "Head of Household". The other individual would receive a new Family ID.
- If an individual moves into an existing family, the individual will be assigned the Family ID of the existing family.
- Family ID's are never "recycled"

In this paper, I assume the household composition does not change immediately after the AVI reform so that I can use the dataset to evaluate households' responses to the reform.

C.1.2 National Level Validation

In this section, I compare the Infogroup data with American Community Survey 5-Year Estimates (ACS) and check the population coverage of the Infogroup dataset. The sample period is from 2010 to 2017.

I use a ratio of population in Infogroup data to the population in ACS data to measure how close the two datasets are. In each year, I calculate the mean and median of the ratio. Table C.1 shows the mean and median over the years.

Table C.1: Yearly Population Coverage of Infogroup v.s. ACS

Year	Mean	Median
2010	0.677	0.694
2011	0.708	0.722
2012	0.795	0.805
2013	0.829	0.839
2014	0.817	0.827
2015	0.890	0.892
2016	0.873	0.879
2017	0.829	0.839

Notes: Table shows the mean and median of the ratio of population in Infogroup data to population in ACS for each year.

The coverage of the Infogroup dataset is fairly good compared with other datasets tracking residential address history. Diamond et al. (2019) use Infutor dataset and characterize the representativeness of Infutor with some validation tests. They link all individuals reported as living in San Francisco in 1990 and 2000 to their census tract to create census tract population counts as measured in Infutor. Then they compare these San Francisco census tract population counts to those reported in the 1990 and 2000 Census for adults 18 years old and above. They find Infutor has a 44 percent sample of the population as the census tract population in 1990 and 110 percent in 2000.

To check whether the Infogroup data disproportionately samples some areas, I regress the sample rate (percentage) on some local characteristic variables in ACS. The covariates are the percentage of white, black, Hispanic, and Asian people, median age, median income (10,000), and the percentage of people with bachelor's degrees or higher. Table C.2 shows that the data covers better in neighborhoods with higher white, black, and Asian populations, lower Hispanic populations, and lower median age. Population coverage is better in neighborhoods with lower median income and lower education levels.

Table C.2: The Bias of Sampling Rate

Dependent Variable:	Population Coverage
White	44.773***
	(1.417)
Black	34.469***
	(1.490)
Hispanic	-26.913***
	(0.662)
Asian	21.806***
	(1.756)
Median Age	-0.0002***
	(0.00000)
Median Income	-0.051***
	(0.003)
Bachelor Degree	-0.00003***
	(0.00000)
Constant	4,359.388***
	(143.448)
Observations	578,046
R^2	0.128
Adjusted R^2	0.128

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. Table shows regression results of the sample rate (percentage) on Census tract level local characteristics. The covariates are the percentage of white, black, Hispanic and Asian people, median age, median income ($10,000), and the percentage of people with bachelor's degree or higher. Local characteristics come from American Community Survey.

C.1.3 Philadelphia Data Validation

Columns 1 and 2 in Table C.3 show the number of households and people living in Philadelphia per year in Infogroup data. Because Infogroup only tracks up to 3 people per household, to validate the population coverage, I calculate the adjusted population of Philadelphia using American Community Survey 1-year estimates (ACS) data. The survey data contain variables showing the number of households of different sizes. For households with 3 and more people, I multiply the household number by 3. For households with fewer than 3 people, I multiply the household number by the household size. Then I aggregate them up to get the "adjusted" population. Column 3 shows the total household number in ACS data and Column 4 shows the "adjusted" population. Infogroup covers fewer populations in 2008 and 2009 and then becomes better in 2010 and 2011. Starting from 2012, it covers more population than that is calculated by ACS estimates.

Table C.3: Household and People in Philadelphia by Year

Year	Household	Population	Household (ACS)	Adjusted Population (ACS)
2006	788652	1030508	554048	1068044
2007	788684	1033350	562384	1095795
2008	671813	915018	578263	1116497
2009	740959	972832	569835	1116732
2010	788091	1023974	575413	1107474
2011	825571	1087293	576429	1113163
2012	899164	1184505	579874	1120474
2013	897711	1235933	582528	1111903
2014	880484	1231890	577862	1117723
2015	943099	1298312	581604	1147407
2016	913886	1296041	580205	1130591
2017	862818	1238513	606142	1126500

Notes: Table shows the total number of household and population in Philadelphia in each year for Infogroup and ACS. Population in both datasets are tracked in a way that households with more than three people are only counted three.

C.1.4 Summary Statistics

The cleaned dataset has 10,692,872 observations (household-year) and 2,102,512 households. Table C.4 shows the summary statistics of household characteristic variables for all the observations.

Table C.4: Summary Statistics of Household Characteristics

Statistic	Mean	St. Dev.	Min	Median	Max
Length of residence (Year)	9.489	10.181	1	6	59
Children count	0.165	0.549	0	0	7
Children indicator	0.130	0.336	0	0	1
Predicted wealth ($1000)	1,617.118	898.938	0	1,531	9,835
Predicted income ($1000)	49.993	51.162	5	35	500
Estimated home value ($1000) 2	101.517	138.691	0	72	9,999
Estimated relative purchasing power ($1000)	47.113	49.156	5	33	500

Notes: Table shows summary statistics of household characteristics in the Infogroup dataset.

C.2 ZTRAX Data

Zillow's Transaction and Assessment Database (ZTRAX) contains two parts, the assessment dataset and the transaction dataset.

The transaction record covers 448,029 unique properties. 38,362 of them do not have parcel ID, which cannot be merged with the assessment record. The rest could be merged with the assessment record with full address and building characteristics. Among the 38,362 records that cannot merge with assessment data, 15,879 of them have address information.

Using the combined assessment/tax data and transaction data, I could create a database showing the address of properties that have a transaction record since around 1995 and the ownership history. For the properties that do not show up in the transaction dataset, I will get the ownership information from the assessment data after 12/30/2018 which shows the owner name of all properties.

In the model estimation, I use ZTRAX transaction information to calculate the average price of each neighborhood in each year.

C.3 CoreLogic

I get the tax and assessment information from CoreLogic data. Philadelphia tax data have tax and assessment information of properties from 2009 to 2017. CoreLogic tax data could be merged with ZTRAX tax data using unique identifiers.

C.4 Construct Philadelphia Datasets

The construction of Philadelphia data contains two steps. In the first step, I extract data from each data source and clean them. In the second step, I combine datasets from multiple data sources together. This section will describe how I clean and merge the data.

C.4.1 Infogroup Data

1. Select all households that have ever lived in Philadelphia from 2006 to 2017.

2. Drop 4 duplicated observations. For each household-year combination, there is only one observation.

3. Fill in gaps.

    - The dataset has a variable "RECENCY_DATE" showing the month and year of the most recent confirmation of this household at this address. If a household is observed in 2017 with a "RECENCY_DATE" earlier than 2017 but is not observed in 2016, I create an observation of this household in 2016 by copying the information in 2017. I repeat this from 2017 back to 2007.

C.4.2 CoreLogic Tax Data

1. Extract the tax information of properties in Philadelphia county.

2. Many properties have two different tax amounts in 2009, so I drop data in 2009.

3. Create a panel data showing the tax and assessment history of properties from 2010 to 2017.

C.4.3 ZTRAX Tax Data

1. ZTRAX data are organized in multiple folders ordered by time. Within each folder, I combine the assessment, tax, address, and building characteristics for all properties in Philadelphia county and create a panel dataset on property-assessment year level.

2. Combine data from all folders. If a property-assessment year combination appears in multiple folders, keep the oldest record.

C.4.4 Ownership Data

I create the ownership history data using the transaction records in the ZTRAX transaction data.

1. Collect the transaction information of all properties in Philadelphia county recorded before January 2, 2020.

2. Keep records that have at least one buyer or seller's name (Records without transactor name are most likely to be mortgage records or foreclosure records.).

3. Create a list of buyers' last names and a list of sellers' last names for each record.

4. Using the lists of buyers' last names, create a dataset showing each property's owners' last names and the owning period from the first transaction record to January 1, 2020, assuming the last buyer owns the property till January 1, 2020.

5. If the first transaction date is later than January 1, 1995, assume the first seller had owned the property since January 1, 1995 and supplement the ownership history.

6. For properties that have never shown up in the transaction data, use the ownership information from December 30, 2018 to April 07, 2020 to supplement, assuming these owners owned the property from 1995 to 2020.

7. Create a panel data on property-year level recording all owners' last names of the property in each year. [1]

C.4.5 Combine Property Level Data

1. Merge the CoreLogic tax data to ZTRAX tax data using Assessment Parcel Number, the unique parcel identifier from the county's assessment office, and assessment year.

2. If a property has a missing address and building characteristics in some years, fill them with the information in other years.

3. Merge the ownership data to the tax data using Assessment Parcel Number and year. If some parcels in the tax data do not have matched Assessment Parcel Number in the ownership data, I try to find matches using the ZTRAX property identifier. If a tax observation matches with multiple ownership observations, combine the owner's last names from the multiple matches.

1. Because the Infogroup data are snapshots of addresses on year level, making ownership data on year level keeps the generality.

C.4.6 Merge to Infogroup Data

I merge the combined property-level data to Infogroup data using property address information from ZTRAX. The merge process contains the following steps.

1. Clean the address information in both datasets. Extract the house number, house number extension, street pre-direction, street name, street suffix, unit type, unit number, 5 digit zip code, and 4 digit zip extension.

2. For each address in Infogroup data, I first find parcels in the property data that have exactly the same house number, street pre-direction, street name, street suffix, and unit number.

3. Then for the addresses in Infogroup data that do not get a match in the previous step and do not have a unit number, I look for parcels that have the same house number, street pre-direction, street name, and street suffix.

4. Among the ones that could match in the previous step, keep the group where Infogroup data do not have a unit number but property data have a unit number. In the next step, I will drop the duplicated matches. In the end, Infogroup addresses without a unit number will be matched with property addresses with a unit number by street address if and only if the match is unique. For matches where Infogroup data have a unit number but property data do not or they have different unit numbers, I will drop them. This ensures that Infogroup unit address will not be matched to property building address and so treat the whole building tax amount as a unit tax amount.

5. Only keep the match pairs where an Infogroup address matches to a unique Assessor Parcel Number.

C.5 Match Quality

C.5.1 Match Rate Across Years

Among 2090449 unmatched observations (households living in Philadelphia in Infogroup data that could not find matches in the property data), 170162 (8.14%) do not have a street address. 3240 (0.15%) do not have a complete address. 111154 (5.32%) only have PO box address. Among the 1805693 unmatched observations with addresses, 652553 (31.22%) cannot be matched even without the unit number. 90597 (4.33%) observations have matches in property data with the same street address, but a different unit number. 568708 (27.21%) observations can get matches in the property data with the same street address, but the property data do not show the unit number. These are usually commercial buildings where the whole building is a single parcel in the property tax data. 494035 (23.63%) observations have more than one match, so they are treated as being unmatched.

Table C.5 shows the number of properties in Philadephia in ZTRAX data and the proportion that could be matched with Infogroup data.

Table C.5: ZTRAX Data Property Count and Match Rate

Assessment Year	Property Count	Matched Property	Match Rate
1999	474901	0	0%
2000	490283	0	0%
2002	490583	0	0%
2003	490978	0	0%
2004	495217	0	0%
2005	498282	0	0%
2006	503179	385137	76.5%
2007	506021	383811	75.8%
2008	507989	345702	68.1%
2009	509811	374073	73.4%
2010	534767	403315	75.4%
2011	535629	408868	76.3%
2012	544015	413878	76.1%
2013	514208	406131	79.0%
2014	542377	405734	74.8%
2015	502350	407417	81.1%
2016	491230	398206	81.1%
2017	580474	406750	70.1%
2018	581960	0	0%

Notes: Table shows the number of properties in Philadephia from ZTRAX data and the proportion that could be matched with Infogroup data.

Table C.6 shows the total number of households living in Philadelphia in Infogroup data per year and the number of households that have tax and assessment information, ownership information, and both information after merging with the other datasets.

Table C.6: Property Information Coverage

Year	Households	With Tax	With Ownership	Both Info	Percentage with Both
2006	788652	0	573098	0	0%
2007	788684	0	573454	0	0%
2008	671813	0	488377	0	0%
2009	740959	0	543689	0	0%
2010	788091	601630	605580	601630	76.3%
2011	825571	631678	635755	631678	76.5%
2012	899164	684073	689092	684073	76.1%
2013	897711	681851	681851	681851	76.0%
2014	880484	673910	675410	673910	76.5%
2015	943099	723892	723892	723892	76.8%
2016	913886	699163	699163	699163	76.5%
2017	862818	672943	691511	672943	78.0%

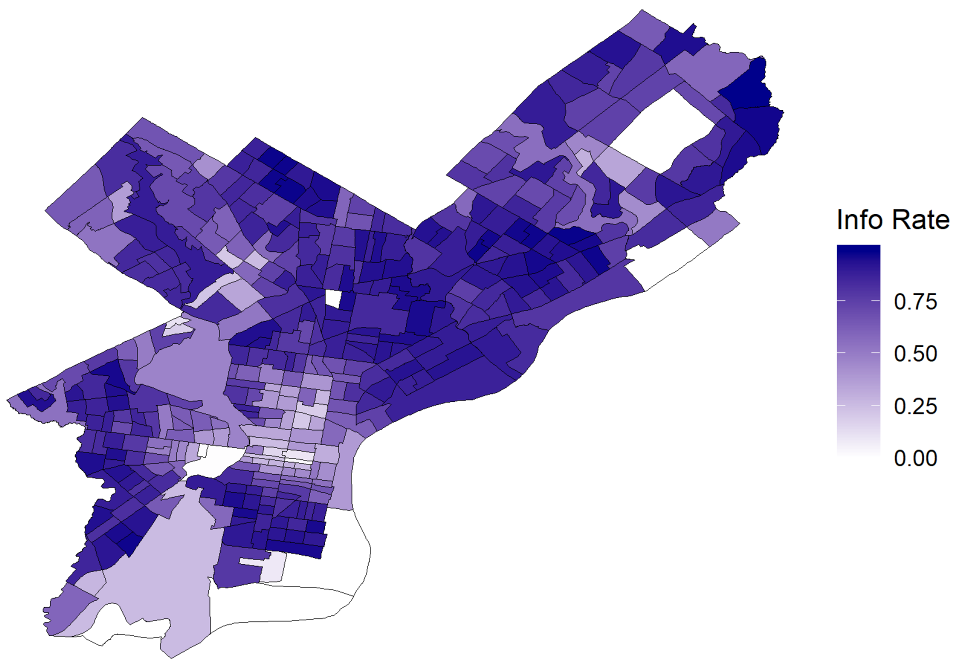
Notes: Table shows the total number of households living in Philadelphia in Infogroup data per year and the number of households that have tax information and ownership information. Tax informaiton comes from Infogroup data and ZTRAX data. Ownership information comes from Infogroup data.

To conclude, almost 80% Infogroup observations could be matched with property data. Around 76% observations have tax, assessment, and ownership information after merge. Around 70% of the mismatches are due to data quality issues. The other 30% are likely to be commercial buildings. The match rate is pretty consistent across years.

C.5.2 Match Rate Across Geographic Areas

I further explore if the proportion of observations that get property information through merge distribute dis-proportionally across Census tracts. Every address in Infogroup data is assigned to a Census tract according the tract boundary delineation in 2010. Figure C.1 shows the average information coverage from 2010 to 2017 across Census tracts.

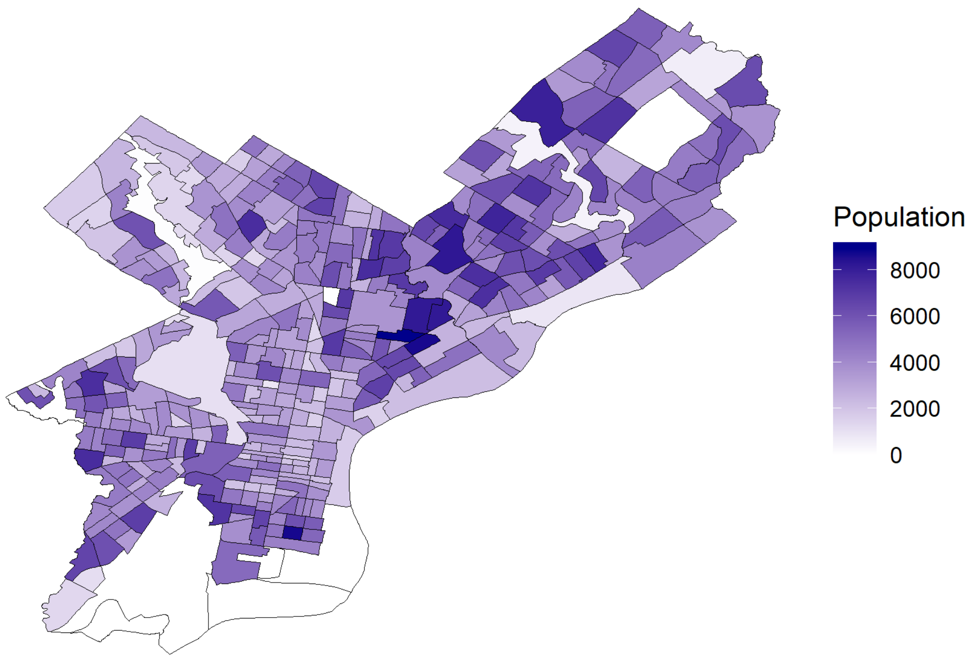
Figure C.1: Property Information Coverage Rate by Census Tracts



Notes: Figure shows the proportion of properties in Infogroup data that have match in each Census tract in Philadelphia from 2010 to 2017.

Through combining the coverage information with American Community Survey (ACS) 5-Year Data, I could explore whether the observations with property information are more concentrated in certain areas. Figure C.2 shows the average population from 2010 to 2017 of all Census tracts.

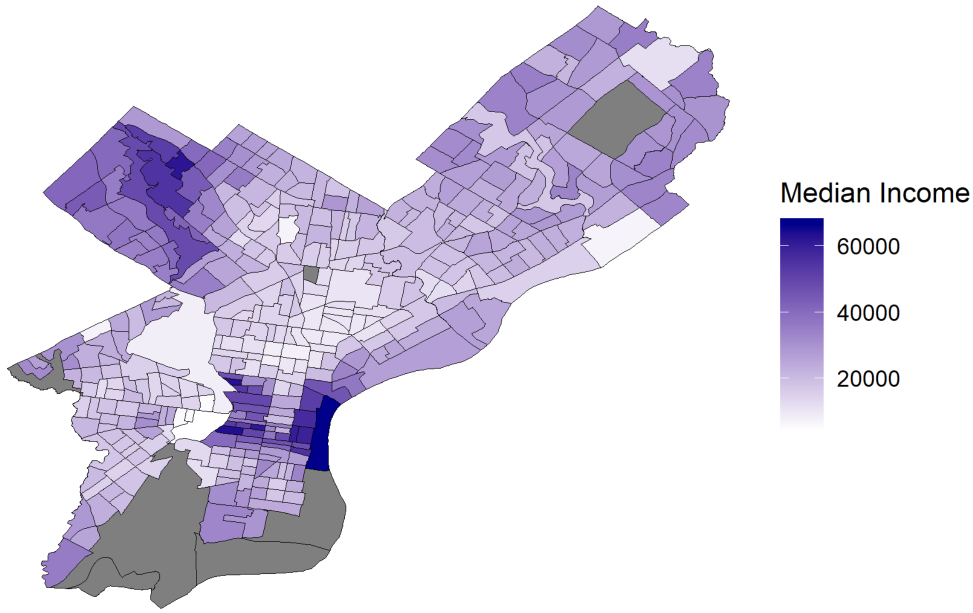
Figure C.2: Average Population (2010-2017) by Census Tracts



Notes: Figure shows the average population of each Census tract in Philadelphia from 2010 to 2017.

Figure C.3 shows the average median income from 2010 to 2017 of all Census tracts.

Figure C.3: Average Median Income (2010-2017) by Census Tracts



Notes: Figure shows the average median income of each Census tract in Philadelphia from 2010 to 2017.

Comparing Figure C.1, C.2, and C.3, I find the match rate is higher in more populated tracts and lower income tracts.

The following correlation table further confirms the above conclusion. The table also shows that the matching rates are higher in regions with smaller proportions of white and Asian people. In regions with more well-educated people, the match rate is lower.

Table C.7: Correlation between Information Coverage Rate and Neighborhood Characteristics

Variable1	Variable2	r	CI_low	CI_high	t	df	p
Info Rate	Population	0.499	0.420	0.571	11.257	382	0
Info Rate	Median Income	-0.152	-0.250	-0.052	-2.986	375	0.114
Info Rate	Median Age	0.056	-0.045	0.156	1.085	375	1
Info Rate	White	-0.167	-0.263	-0.067	-3.283	376	0.044
Info Rate	Black	0.145	0.045	0.242	2.837	376	0.173
Info Rate	Indian	-0.075	-0.175	0.026	-1.461	376	1
Info Rate	Asian	-0.176	-0.272	-0.077	-3.476	376	0.023
Info Rate	Pacific	0.009	-0.092	0.110	0.169	376	1
Info Rate	Less Than High School	0.244	0.147	0.337	4.881	376	0.0001
Info Rate	High School	0.494	0.414	0.567	11.015	376	0
Info Rate	College	0.354	0.263	0.439	7.345	376	0
Info Rate	Bachelor	-0.471	-0.546	-0.389	-10.352	376	0
Info Rate	Graduate	-0.505	-0.576	-0.425	-11.330	376	0

Notes: Table shows the correlation between Infogroup data match rate for each Census tract and tract characteristics.

I also checked whether the geographical coverage pattern changed over time. I did not find drastic change over time, especially around 2014, when the AVI reform took place.