Aalto University
School of Business

THE CAPITALIZATION OF MUNICIPAL INCOME TAXES ON HOUSE PRICES

Evidence from the Finnish housing markets

Master's Thesis
Niko Peltola
Aalto University School of Business
Finance
Fall 2021

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Niko Peltola	
Title of thesis The capitalization of municipal income taxes on house prices	
Degree Master of Science (Economics and Business Administration)	
Degree programme Master's Programme in Finance	
Thesis advisor(s) Michael Ungeheuer	
Year of approval 2021	**Number of pages** 55 **Language** English

Abstract

Finnish municipalities have relatively high level of independence in setting their tax rates, and differences between neighboring municipalities can be significant. In theory, the present value of tax savings in a low-tax area versus high-tax area should be perfectly capitalized on house prices. Any deviation from this would be a sign of market mispricing.

This thesis is the first to study the capitalization effect in Finland. It also contributes to the existing literature with one of the most high-quality datasets used. Furthermore, I can to some extent mitigate the issue of differences in public service quality, which reduces the validity of several earlier studies.

My empirical design is based on spatial differencing. The main identification strategy utilizes the areas nearby municipal borderlines and differencing the closest matching dwelling features across the border. This makes it possible to control for the location of the dwelling robustly and with little assumptions. I conduct several robustness checks, including subsets of municipalities, dwelling types and price classes. Methodologically, I further test the effect with border fixed effects and a gradient boosting pricing model, as well as a distorted placebo dataset.

I argue that municipal income taxes are undercapitalized on house prices. My preferred estimate is that around half of the tax differences are capitalized with an appropriate discount rate. However, there is variation in the level of capitalization between municipalities, dwelling types and dwelling price classes. I find that the most expensive dwellings in expensive regions drive the observed capitalization, whereas in less expensive dwellings the capitalization is close to zero. The results suggest that motivation and skill are required from the house buyers to exploit the tax differences.

Keywords tax capitalization, municipal taxation, housing market, spatial differencing, boundary discontinuity

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Niko Peltola

Työn nimi Kunnallisen tuloveron kapitalisoituminen asuntojen hintoihin

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Rahoitus

Työn ohjaaja(t) Michael Ungeheuer

Hyväksymisvuosi 2021 **Sivumäärä** 55 **Kieli** englanti

Tiivistelmä

Suomalaiset kunnat voivat verrattain itsenäisesti päättää kunnallisverostaan, ja erot myös naapurikuntien välillä voivat olla merkittäviä. Teoriassa korkean ja matalan verotuksen alueiden alueilla sijaitsevien asuntojen vuotuisten verojen nykyarvojen tulisi kapitalisoitua täysin asuntojen hintoihin. Tästä havaitut poikkeamat kertovat markkinoiden hinnoitteluvirheestä.

Tässä tutkielmassa tutkitaan ensimmäistä kertaa kunnallisverojen kapitalisoitumista Suomessa. Tutkimus kontribuoi aiheesta nykyisin olemassa olevaan tutkimukseen myös käytettävän datan määrän ja laadun merkittävällä parantamisella. Tämän lisäksi tutkimuksessa huomioidaan julkisten palveluiden laatuun, joka heikentää monien aiempien tutkimusten luotettavuutta.

Empiirinen strategiani perustuu *spatial differencing* -menetelmään. Strategiassa hyödynnetään kuntarajojen lähellä sijaitsevia asuntoja, muodostaen lähellä toisistaan sijaitsevista asunnoista pareja ja tutkimalla parien ominaisuuksien erotuksia. Tämä mahdollistaa sijainnin hintavaikutuksen kontrolloinnin luotettavasti ja vähäisin oletuksin. Arvioin kapitalisoitumista myös rajatummin otoksin eri kunnista, asuntotyypeistä ja hintaluokista. Lisäksi käytän vaihtoehtoisia metodeita, kuntarajojen kiinteitä vaikutuksia, *gradient boosting* -hinnoittelumallia sekä placebodataa.

Tulosten mukaan kunnalliset tuloverot eivät täysin kapitalisoidu asuntojen hintoihin. Arvioni mukaan kapitalisaation määrä on noin puolet verojen todellisesta nykyarvosta käyvin diskonttokoroin. On kuitenkin huomioitava, että kapitalisaation asteessa on eroja kuntien, asuntotyyppien ja hintaluokkien välillä. Tulosteni mukaan kapitalisaatiota tapahtuu erityisesti kalliimmissa asunnoissa ja kalliimmilla alueilla. Edullisimmissa asunnoissa kapitalisaatio on lähellä nollaa. Tulokset viittaavat siihen, että asunnon ostajien motivaatiolla ja taidoilla on vaikutusta verojen hintavaikutuksen huomioimiseen ja hyödyntämiseen.

Avainsanat verokapitalisaatio, kuntaverotus, asuntomarkkinat, *spatial differencing, boundary discontinuity*

Table of Contents

List of figures and tables

1. Introduction

1.1. Research overview and motivation

This thesis studies the capitalization of municipal income tax rate differences on dwelling prices in Finland using a quasi-natural empirical design. In theory, any differences in the tax rate between two dwellings in different municipalities should be fully reflected on the dwelling price with an appropriate discount rate. Any deviation from this is a sign of market mispricing.

Evidence of potential pricing errors in the housing market has economic and social significance. For most people, buying a dwelling is the biggest single investment of their lives. In Finland, people's own homes form 48% of the total household wealth, and 65% of the households live in dwellings owned by themselves. (Statistics Finland, 2021). Better understanding of the pricing factors and improving the accuracy of pricing real estate makes the housing market more efficient, liquid, and equal for all market participants. Furthermore, the capitalization of taxation provides an interesting signal of people's sensitivity to changes in municipal taxes, which could be valuable for decision makers.

I study the tax capitalization by using transaction-level data of dwellings[1] sold in Finland between 2000 and 2021. My methodological approach is based on geographical border discontinuity: studying the border regions between municipalities to control for the effect of location and using the tax differences as a treatment variable to show causal evidence of the tax capitalization on dwelling prices.

The results indicate that municipal income taxes do capitalize on dwelling prices, but only partially. Different specifications suggest a capitalization rate of around roughly 50 percent, calculated as the ratio of quality-controlled dwelling price differences and present value of the tax differences. I also find that there is significant variation in the level of capitalization between geographical regions and house price classes, suggesting that different groups of house buyers have different motivation, skills, or market ability to take taxation into account when buying housing.

[1] The dataset contains various residential building types, mainly single-family homes and dwellings in apartment buildings and row houses. For simplicity, I refer to all of these units as *dwellings* in this paper.

1.2. Behavioural background

In the literature of behavioural finance and economics, most of the asset valuation experiments are based on the stock markets. Compared to stocks, housing markets are more illiquid, contain asymmetric information and most of the market participants are unsophisticated investors. For non-professional dwelling buyers, properties are hard to valuate as there is little reference data publicly available and the buyers are generally very inexperienced in the valuation task. Unlike the buyers, the sellers are typically backed by professional realtors, with access to transactions data and experience in the valuation. This creates a market setting with highly asymmetric information. Finally, dwelling as an investment also often creates a significant emotional attachment to both the seller and buyer. All of these factors are shown to increase misjudgements and biased behaviour among people. (Hirshleifer, 2001)

Studying people's economical behaviour and potential "bounded rationality" in the housing market context is interesting for several reasons. First, bidding for a dwelling raises a classic example of the *winner's curse*. That is, the one buyer that is ready to offer the highest bid sets the market price for the house. Therefore, for undercapitalization to occur, it is enough if that one buyer does not fully capitalize municipal taxes in their bid. Second, the illiquid and non-securitized nature of housing markets mean that if pricing errors exist, it is difficult for anyone to exploit the errors in search of arbitrage returns by for example short-selling assets. Therefore, unlike generally in stock markets, pricing anomalies can be substantial and also persistent.

On the individuals' level, dwelling buyers are often severely underinformed of the pricing factors of a house, and dependent on heuristic approaches of price formation. Furthermore, people are often biased towards hyperbolic discounting, meaning the overweighting of present and underweighting of future cash flows. (Elster, 1979). So, even if the dwelling buyers were aware of the tax rate differences, the discount rate for the cash flows could be unreasonably high. Based on this background, I hypothesize that the municipal taxes are unlikely to fully capitalize on housing prices.

Due to the behavioural elements of the capitalization, I also hypothesize that higher income might be correlated with higher level of tax capitalization. House buyers with higher income are typically more educated and might therefore have the capacity and skills to take the tax factor into account and discount it with an appropriate rate. Higher income also naturally makes the value of the tax savings higher in absolute terms, increasing the motivation of the buyer to pay attention to the taxes.

1.3. A case example

As an illustrative case example of the decision-making problem, let's think of two identical dwellings A and B at different sides of the border of Helsinki and Vantaa. The tax rates of these neighbouring municipalities in 2021 are 18.0% and 19.0%, respectively. Let's think of a household of two searching for an apartment in the region. Both members of the household have an average Finnish annual gross salary of €40,000. With this income level, a 1.0 percentage point difference in the tax rate equals a 0.85 percentage point difference in the effective tax rate. This is equivalent to an annual difference of €656 in taxes paid for the household.

If the applicable discount rate is 2% and the taxes are paid until perpetuity, the present value of these taxes would be €32,800. The average dwelling price within a 1-kilometer radius of the border of Helsinki and Vantaa is €193,000 which means that the tax effect would in this case cause a 17 percent price difference. If the case example couple is otherwise indifferent in choosing between A and B, this should be the price difference the dwellings. Equivalently, the observed price difference of dwellings will reveal the implied discount rate of the home buyers.

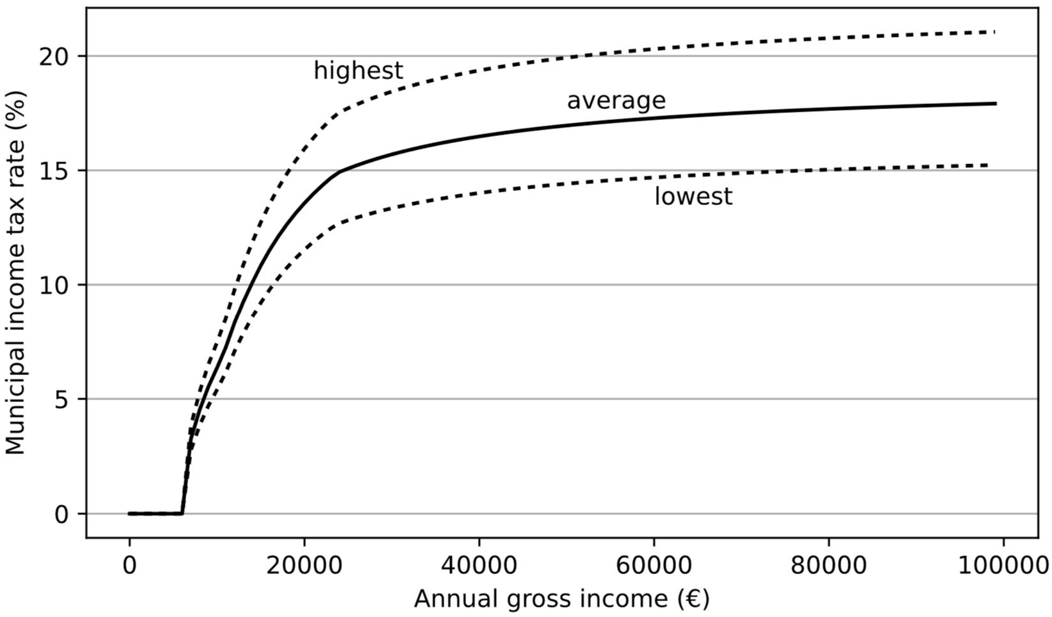
This case example shows the magnitude of the economic significance of the tax capitalization. The border of Helsinki and Vantaa is relatively densely populated, and as shown, the present value of the taxes paid varies significantly even for average income citizens. Therefore, evidence of how these differences are capitalized on dwelling prices has practical significance.

1.4. Municipal finance and taxation in Finland

In Finland, income taxation comprises of taxes for three receivers: the state, municipalities, and churches. Out of these taxes, the state taxation regime is the same for all citizens. The church tax is only paid by those who are a member of either of the two state churches. However, the municipal tax is dependent on the income earners place of residence and is impliedly chosen by a citizen when they buy a house. Unlike the progressive state tax, the municipal tax rate is flat. However, as illustrated in Figure 1, various deductions make the effective tax rate progressive as well, especially on lower income.

Figure 1: Effective municipal tax rate on different income levels

Figure 1 shows the dependency of municipal tax rate and annual gross income. The effective tax rates are calculated using the standard tax deductions that every taxpayer is entitled to. The solid line represents the average tax rate (20.0%), and the dashed lines the highest (Halsua, 23.5%) and lowest (Kauniainen, 17.0%) tax rates in mainland Finland in 2021. Data source: Taxpayers Association of Finland.



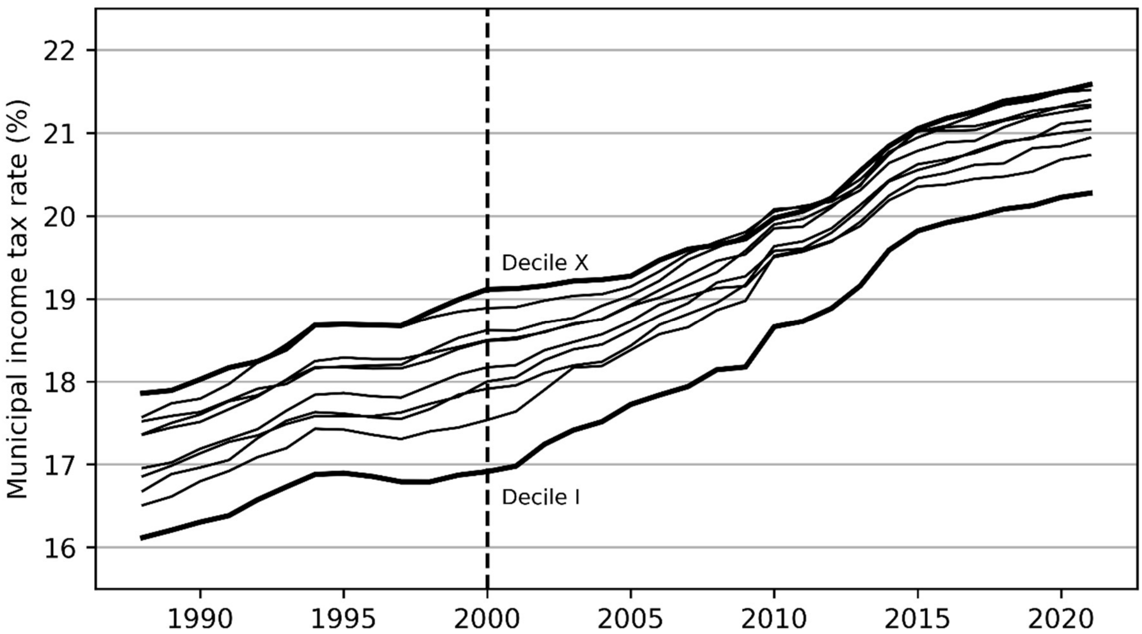
Municipalities may choose their tax rates independently. In 2021, the rate ranges between 17.0% – 23.5%, the average being 20.0%. The average tax rates have been rising during the past decades, driven by aging population and the increased number of mandatory municipal services. (Viherkenttä, 2021) Also, the state has increased the deduction for work income, increasing the progression of the income tax and forcing municipalities to increase their tax rates. (Holm & Huovari, 2011)

In the tax capitalization setting, the consistency of the tax rates is an important assumption, as the present value of the future taxes is calculated using the tax rate of the sales time. The dwelling buyer may estimate that the level of taxation might increase overall, but it is more difficult to estimate the predict the tax rate development with relation to neighbouring municipalities. Furthermore, an argument can be made that the tax rates converge towards a mean in long term, diminishing the differences between municipalities and then also the present value of any tax savings.

Figure 2 illustrates both the development of the average tax rates as well as the persistency of the levels of the tax rates within municipalities. The municipalities are divided into deciles based on their income tax rates on year 2000, as that is the first year of observations in my dwelling sales dataset. The figure shows that the tax rates have overall risen significantly in the past 30 years. Also, the deciles behave similarly with relation to each other, indicating that the levels of the tax rates are fairly persistent within municipalities. Some convergence of the tax rates can be observed, but this may be mostly due to mechanical mean reversion. I analyse the potential tax converge further in Appendix A3 and find that while some convergence might exist, it should not critically violate the assumptions of my empirical design.

Figure 2: Development of municipal income tax rates by municipality deciles

Figure 2 shows the development of the municipal income tax rate in Finland during 1988-2021. The data are divided into deciles based on the municipalities' income tax rates on year 2000, deciles I and X containing the lowest and highest tax rates, respectively. The bolded lines represent the I and X deciles, and the thin lines the II-IX deciles. Data source: Association of Finnish Municipalities.



Finnish municipalities vary a lot in terms of citizen age structure, population density and share of foreigners. To level the service requirement imbalances and fiscal deficits, the Finnish state grants intergovernmental transfers to municipalities. These grants are a significant source of service financing for most of the rural municipalities in Finland. In 2019, 266 municipalities received in total €1.4 billion in grants, whereas 29 net payer municipalities paid €650 million. The difference is paid by the Finnish state. The grant calculation formula is not dependent on the financial performance of single municipalities, i.e. the municipalities are not penalised for implementing cost savings or increasing their tax revenue. (Salonen & Nissinen, 2019)

The state policy also aims to diminish tax competition between municipalities. This kind of competition is seen as harmful as it may increase the inequality between regions, create inconsistency in the taxation regime and risk the financing of municipal services. The ability to compete with taxes has been reduced by for example setting a minimum property tax rate. Also, the fact that municipal income tax must be set as a flat rate makes it impossible for municipalities to lure wealthy individuals with tax discounts. (Peni, Riipinen, Tukiainen, & Viitala, 2021). Analysis on property taxes shows that the policies are efficient: contrary to what has been observed in many other countries, there is no evidence of tax competition between Finnish municipalities. (Lyytikäinen, 2012).

2. Literature review and contribution

2.1. Evidence on tax capitalization

The question of the rate of tax capitalization on property prices is rather classical in real estate economics and bases on the article by Oates (1969). He concludes that property tax differences between cities in New Jersey were almost fully capitalized on dwelling prices, as the theory suggests. Following that paper, mixed evidence has been documented on the level of the capitalization. In addition to Oates (1969), for example Reinhard (1981), Palmon & Smith (1998), Gallagher, Kurban, & Persky (2013), Borge & Rattsø (2014) and Høj, Jørgensen, & Schou (2018) also find evidence supporting close to full capitalization. Other papers report significantly lower levels or even zero capitalization of property taxes. (Charlot, Paty, & Visalli, 2011; Follain & Malpezzi, 1981; McMillan & Carlson, 1977).

In addition to cross-sectional differences in tax rates, various policy changes have been used as exogenous shocks to show causal evidence of tax capitalization. Høj et al. (2018) study a Danish land tax reform in 2007 that caused varying changes in the tax rates in different municipalities. They find that the tax changes were fully capitalized to dwelling prices with a reasonable discount rate of 2.3 percent.

Hipsman (2018) is one of the few to study the capitalization of specifically income taxes. He uses an aggregated US dataset of dwelling price indices to study the effect of changing state income taxes on home values. He finds "strongly suggestive but not conclusive evidence that home prices are responsive to state income taxes". His results suggest that dwelling prices are indeed negatively affected by increased state income taxes. Furthermore, he finds that the price impact is capitalized on dwelling prices as early as two years prior to the tax change. This could be due to buyers anticipating the tax change or due to exogenous economic conditions that drive dwelling prices and taxation decisions. However, the evidence from Hipsman (2018) is far from conclusive. The results are not consistent over different specifications and suffer from large standard errors. He proposes further research to be done with transaction level data and a regression discontinuity design.

In Finland, Kauko (2002) studies a neural network based spatial dwelling pricing model using Finnish data and also employs municipal taxes in the model. His model design does not provide any causal evidence of the existence or level of tax capitalization. However, he presents the relative importance of the variables used in the model and finds that taxation is among the least important factors, while factors like services and distance to city centre drive prices the most. It should be noted that his result does not make any claim on the level of tax capitalization, just that statistically the tax factor is diminished by other factors. Kauko (2002) also conducts an expert interview and a survey and similarly finds that taxes are on average not seen as important factor in dwelling valuation.

My study contributes to the existing literature by improving the empirical setting of previous studies on local tax capitalization, as well as mitigating the issue of controlling for differences in public service quality. It is also the first study of the capitalization effect in Finland. I next discuss the contribution of this study in more detail.

2.2. Empirical contribution

First, most of the previous literature analyses the capitalization of property or land taxes, whereas my focus will be in the municipal income tax. The income tax is more relevant in Finland, as it provides a lot more variation than property tax: the annual municipal income taxes for a median income couple range between €11,400 and €15,800 while the property taxes for a median priced dwelling only varies between €200 and €450 in different municipalities. (Kalluinen, 2021) The income tax is also simpler to calculate, as Finnish property taxes are not based on the market values of dwellings, but instead an estimate of the building costs. Also, the income tax only concerns natural persons, unlike property and land taxes.

Second, the available dataset is significantly richer than in reference studies. With over 1.2 million records of sales transactions in total and around 100,000 with a less than 1 kilometre distance to a municipal border, I have by a magnitude more datapoints than most of the previous studies with related topics. Due to the private nature of real estate data and costs for acquisition, most papers suffer from lack of quality data. The sample sizes for transaction level data have been growing in recent papers but are still relatively small: 501 records (Palmon & Smith, 1998), 10,000 records (Charlot et al., 2011), 60,000 records (Giertz, Ramezani, & Beron, 2021). Even methodological studies seem to suffer from small datasets; for example Helbich, Brunauer, Vaz, & Nijkamp (2013) study a related spatial regression methodology with a sample of 1,700 records. Recent Nordic papers have been able to use somewhat larger datasets: 73,000 (Borge & Rattsø, 2014) and 216,000 dwelling sales (Høj et al., 2018). However, these samples are not restricted to the borders, but contain all locations.

Some papers use aggregated data instead of transactions: for example, Oates (1969) uses aggregate data from 53 municipalities in New Jersey, and McMillan & Carlson (1977) from 65 rural towns in Wisconsin. Hipsman (2018) uses average prices aggregated on ZIP code level. The robustness of his evidence suffers from inconsistent empirical results, which could be driven by the challenges caused by the aggregation. All in all, aggregated datasets cannot provide the same level of control for differing area characteristics and dwelling features than transaction-level data.

Third, Finland as a region provides excellent ground for analysing tax differences. As there are around 900 borders between the 310 municipalities of Finland, the data provides enough variation to study the effect of cross-border differences. (See *Figure AX* in Appendix). Out of the 900 borders, 474 borders contain dwellings within a 1-kilometre radius in my dataset. This is significantly more than for example the 109 borders between the 50 US states. Furthermore, as most of Finland is relatively sparsely populated, the dwellings are to some extent spread around the regions and not only clustered in cities. I find that around 10% of the dwellings in my sample are located within 1 kilometre distance of an inner municipal border. I discuss the features of the data more extensively in *Section 3*.

Finally, a major contribution of this paper is that it allows to study the tax capitalization in an environment where taxes are not highly connected with public service quality. The connection is a challenge in the empirical design of most of the earlier studies. I next discuss the phenomena in more detail and how my research setting mitigates the endogeneity issue.

2.3. Taxes and public services

2.3.1. Public services in tax capitalization literature

A typical issue raised in papers studying tax capitalization, especially in the U.S, is the interconnectedness of taxes and public spending. The theoretical background of this effect is formalized by Tiebout (1956), who states that each independent region chooses their optimal balance level of taxation and public services, and people may choose the region that has the best balance for them. In the perfect Tiebout equilibrium, higher taxation means higher level of public services, and this offset would make the tax level itself irrelevant for dwelling prices. If there is a preferrable package of taxes and services in some region, people can vote with their feet and move there. This should be reflected in housing prices as well.

Palmon & Smith (1998) show that some of the earlier evidence supporting limited tax capitalization can be explained by failing to control the public services aspect well enough. They therefore suggest that taxes are in fact more capitalized than what the majority of the literature proposes.

The Tiebout mechanism has been studied a lot from the perspective of school quality differences. In countries where pupils attend standardized tests and school results are publicly available, the data can be used to find how school quality is capitalized on dwelling prices. Discrete school catchment areas make it possible to apply regression discontinuity methods for causal evidence. A majority of the literature proposes that higher school quality is clearly capitalized on dwelling prices. (Bayer, Ferreira, & McMillan, 2007; Black, 1999; Black & Machin, 2011). Finnish evidence proposes that significant capitalization exists even if school quality information is not public and the differences are small. (Harjunen, Kortelainen, & Saarimaa, 2018)

Gallagher et al. (2013) mitigate the school quality issue by studying property taxes that follow school district boundaries in Cook County, Illinois. As the district boundaries only affect the quality of schools, they study capitalization on small single-person houses with no school-aged children. This allows them to estimate the level of tax capitalization free of public service interference.

2.3.2. Research setting with Finnish public services

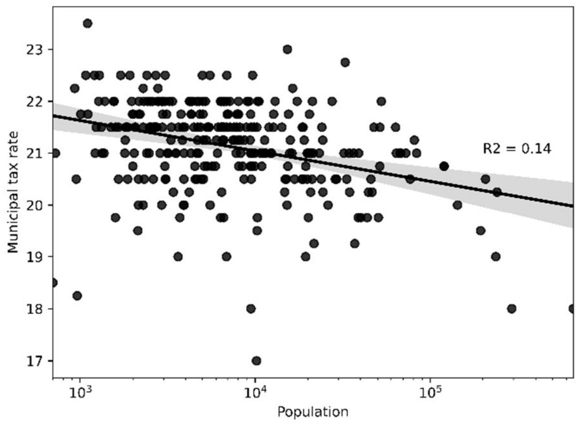
I argue that my research setting has a unique advantage in separating taxes and public services. This is due to the fact that the duties of Finnish municipalities are precisely set in the Local Government Act. For example, citizens are entitled to the same medical care services regardless of the place of their residence. Neighbouring municipalities typically work in contractual co-operation to provide health care services for their citizens, who can choose where they use the services even across municipality borders. (Heinämäki, 2011)

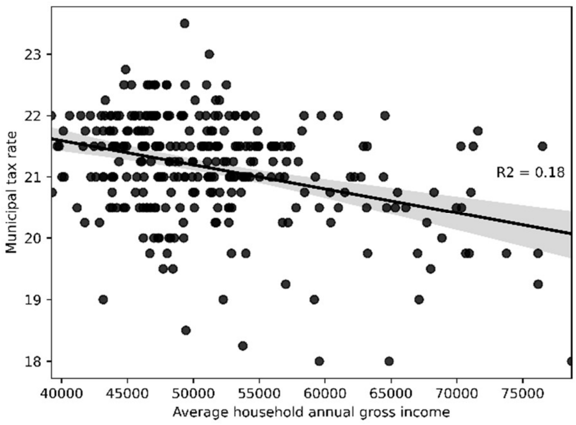
Also the schooling services are relatively equal in municipalities across the country. As an example, Finland had the smallest differences in the latest PISA test scores between schools among the 80 participating countries. (Ministry of Education and Culture, 2019) Preventing socio-economic segregation is a stated and important objective in schooling and housing policies. For example, schools that have a lot of pupils with lower socio-economic or immigrant background receive more public funding. (Harjunen et al., 2018). What is noteworthy is that in my research design, the impact of the dwelling location is controlled by restricting the sample to municipal borders and matching the dwellings with their closest neighbours. Even if schools' learning outcomes differ due to socio-economic selection between neighbourhoods, the neighbourhood quality should be approximately same when the dwellings are nearby each other, just at the different sides of the border.

My counterargument is that the tax rates of Finnish municipalities are more driven by the demographic composition and income level of the citizens as well as public sector efficiency and economies of scale, rather than different quality of public services between municipalities. Figure 3 shows that the tax rate is negatively correlated with both municipality size and the average household gross income of the citizens.

Figure 3: Population and household income correlation with municipal tax rates

Figure 3 shows the correlation between municipal tax rates and 1) population; 2) the average household annual gross income in the municipality. The data contains the individual municipalities. Data source: Statistics Finland.



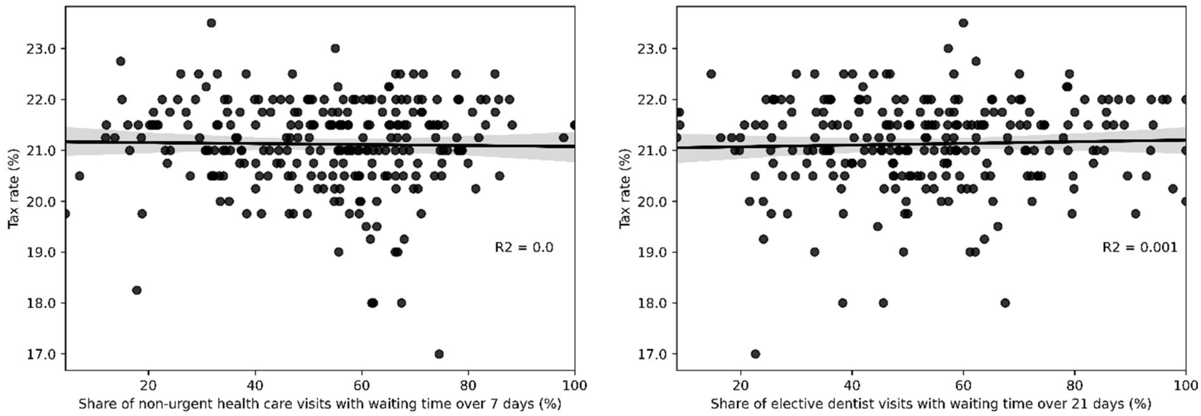


If municipalities were to variate their service and tax level in line with the Tiebout equilibrium, this should be reflected in the measures of service quality. In Figure 4a, I use the share of long waiting times to public health care visits as a proxy for the municipal service quality. In Figure 4b, I report aggregated survey results where citizens are asked about their satisfaction to certain municipal services. I regress both of these measures against the tax rate and find no co-variation between the variables.

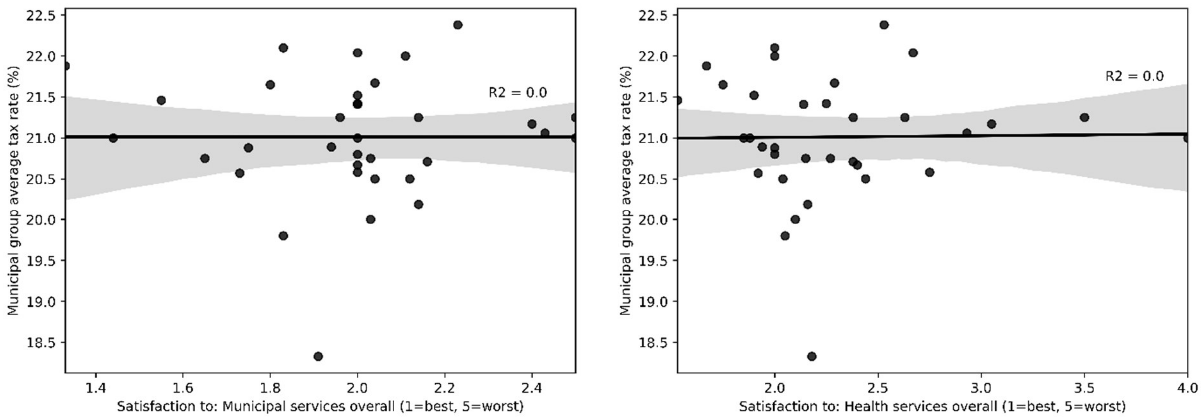
Figure 4: Municipal service quality and tax rate

Figure 4 shows the correlation between municipal tax rates and A) share of healthcare patients with long waiting times; B) surveyed satisfaction to overall services and health services. The data in Panel B is aggregated by county and municipality size, due to data availability. Data sources: Statistics Finland, Sotkanet, Aila.

Panel A: Share of healthcare patients with long waiting times (over 7 / over 21 days)



Panel B: Satisfaction to municipal services (survey results)



Any conclusive evidence of the connectedness of municipal taxes and service quality is difficult to draw. However, I claim that my research design and data provide a significant advantage compared to previous literature in assuming the exogeneity of the factors.

3. Data

3.1. Dwelling transaction data

The dwelling sales data are provided by the Finnish Federation of Real Estate Agency (KVKL). The collected dataset includes the dwellings sold by the major real estate agencies in Finland during 1/2000 – 6/2021 and has in total 1.2M records of sales. All in all, the dataset contains the records of around 80% of all dwelling sales in Finland and should be in all measures a representative sample of the housing market. The data is available upon request for research purposes, but due to its confidentiality, any unit-level details or visualizations will not be reported in this paper.

I clean the data by removing all non-residential properties and properties outside mainland Finland (mainly Åland islands). As each record has been manually added by a human realtor, I study each variable for clear errors, outliers and typos and remove such records. The data cleaning process is described in Appendix A4.

As my empirical design focuses on dwellings nearby municipal borders (see *Section 4.2.*), it is critical that enough dwellings fulfil that condition. I find that around 10% of the dwellings are located within 1 kilometre distance of some municipal border. The final dataset with the 1-kilometre threshold then contains 100,384 dwelling sales records. As discussed, this is significantly more than in most of the reference studies of tax capitalization.

3.2. Variable selection

The selection of the dwelling characteristics control variables in is based on earlier literature with the same dataset (e.g., Harjunen et al., 2018; Mäenpää & Mäenpää, 2021; Rantanen, 2019), as well as earlier tax capitalization research. (e.g., Charlot et al., 2011). The variable specification is summarized in Table 1*:*

Table 1: Variable selection and calculation

Table 1 presents the selection, units and calculation methods of the variables used in all of the empirical models in this paper, unless stated otherwise.

	Variable	Unit	Calculation
Dependent variable	Price / m^2	€/m^2	Price / Living area
Treatment variables	Tax (%)	float	Municipal income tax rate as a floating number (e.g., 19.5)
	Tax (€1000 per annum)	€	Estimate of the annual municipal income tax payment for the household. Calculation specified in *Section 4.3.*
Control variables	Living area	m^2	Logarithmic
	Age	years	Sales year – Building year
	Apartment x Age > 55	binary	Dummy variable, 1 if the building type is an apartment and age > 55. Represents the estimated time of a first plumbing renovation.
	Condition	category	Fixed effects: 'poor' (reference), 'satisfactory', 'good', 'new'
	Own lot	binary	Dummy variable, 1 if the building is on free-hold lot, 0 if on rental lot
	Floor	number	Floor number of dwellings in apartment buildings, 0 for other building types
	Sales year	year	*Spatial differencing:* matching criteria *Boundary fixed effects:* fixed effect

3.3. Geospatial data processing

To find the municipal boundary lines, I first scrape the all the pairwise neighbouring municipalities from the website of Association of Finnish Municipalities (Kuntaliitto)[1]. Next, I collect the municipality borders as of 2021 as a shapefile from the National Land Survey of Finland (Maanmittauslaitos).[2] I process the shapefile by going through the list of the collected neighbouring municipality pairs and finding the common coordinates for each of them. This

[1] https://www.localfinland.fi/municipalities/
[2] https://tiedostopalvelu.maanmittauslaitos.fi/tp/kartta

process yields a list of the borderlines as coordinates, labelled with corresponding municipality pairs.

As the next step I map each dwelling to the closest of the municipal borderlines. As the KVKL data includes the location of the dwellings as street addresses, I need to transform the information into coordinates. I implement the geocoding with the *Addresses, postal codes and WGS84-coordinates of Finnish buildings* open dataset provided by the Digital and Population Data Services Agency (DVV).[1] After cleaning the addresses information in the KVKL data, I merge the datasets and find matches to 93% of the records. The unmatched addresses are mainly written in some unsystematic way, spelled wrong, or the street name has changed.

Finally, I calculate the closest distance of each dwelling and each borderline in the municipality of the dwelling location. I map the dwellings to the closest borderlines and restrict the baseline sample to the dwellings that have a distance of less than 1,000 meters from the closest border.

3.4. Other data sources

The municipal tax rates are privately provided by the Association of Finnish Municipalities (Kuntaliitto). The panel data includes the income tax rates for all municipalities for the time period of 1988-2021.

I also collect information of municipal services and public spending from the Finnish Institute of Health and Wellbeing (THL) open database Sotkanet[2] as well as the Finnish Social Science Data Archive open database Aila.[3]

[1] https://www.avoindata.fi/data/fi/dataset/postcodes
[2] https://sotkanet.fi/sotkanet/
[3] https://services.fsd.tuni.fi/

4. Methods

4.1. Hedonic pricing model

To study the effect of taxes on dwelling prices, a pricing model is needed to standardise the effect of other relevant features of the house. The standard hedonic pricing function of a dwelling in the tax capitalization setting is as follows:

$$p = X'\beta + z\gamma + t\delta + \varepsilon,$$

where:

p = square-meter price of the dwelling,

X = vector of dwelling/transaction features, such as living area and condition,

z = unobservable location attributes, like generic area desirability, and

t = the rate of municipality tax in the dwelling location, and

ε = the error term.

However, the key challenge in analysing the effect of municipal tax rates is the fact that it is tied to the location of the dwelling and therefore to the unobservable location attributes, i.e. $\mathrm{Cov}(\mathrm{z}, \mathrm{t}) \neq 0$. Due to the nature of real estate as an asset, location is a strong driver of dwelling prices and must be carefully addressed in the research setting.

4.2. Geographic regression discontinuity

To solve the omitted variable bias regarding location, various methodological strategies can be used. Generally, using a *regression discontinuity* (RD) design makes it possible to show causal evidence with weak assumptions. As introduced by Cushing (1984), a growing number of real estate studies use a *border discontinuity* design that allows separating the unobservable features from the one under research by spatial identification. This design is based on the assumption that the unobservable features (like neighbourhood desirability) vary spatially smoothly, and the arbitrarily placed border causes a discrete change only on the variable of interest. (Keele & Titiunik, 2015)

4.2.1. Spatial differencing

My baseline model to implement the capitalization analysis utilizing the municipal borders is *spatial differencing* (Gibbons, Machin, & Silva, 2013; Harjunen et al., 2018). In practise, every dwelling nearby a municipality border is matched with its closest neighbour on the other side of the border. A hedonic pricing model is then used to analyse the price difference of these units that is explained by the municipal tax difference. In my research setting where municipal borders are long and there are a limited number of observations at each border, pooling all the boundary dwellings together weakens the assumption of similar location preferability. Black & Machin (2011) show that the spatial matching identification method should provide the most accurate results in these kind of research settings.

The spatial differencing model can be derived from the described standard hedonic pricing function by studying the differences of the spatially matched dwellings *i* and *j*:

$$p_i - p_j = (X_i' - X_j')\beta + (z_i - z_j)\gamma + (t_i - t_j)\delta + (\varepsilon_i - \varepsilon_j)$$

As the dwellings are matched with a nearby neighbour, the unobserved location attributes z_i and z_j are equal, and the term is cancelled out. Also, the errors ε_i and ε_j are assumed uncorrelated. The final model only consists of the measurable differences of the dwelling prices, features, and taxes:

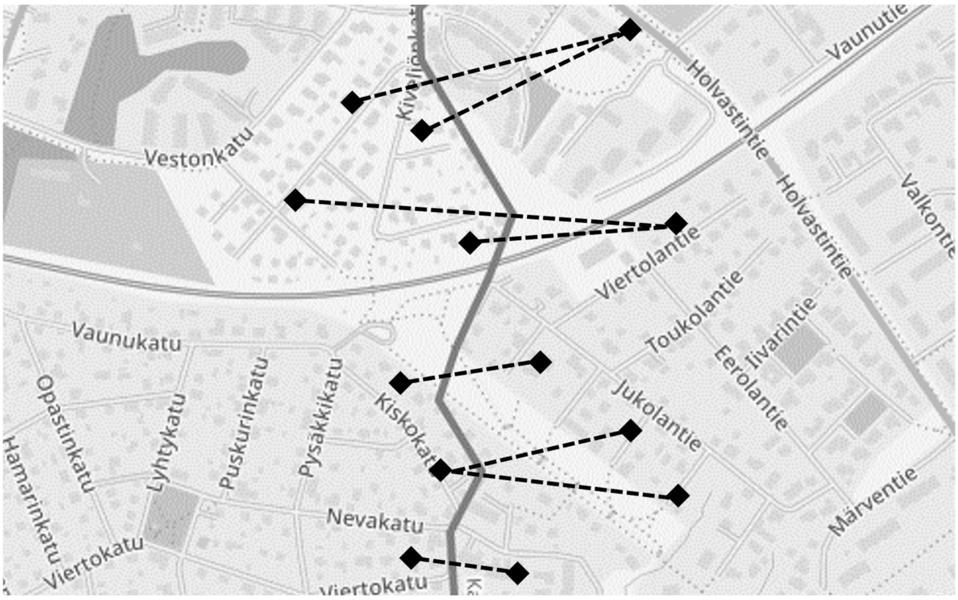
$$p_i - p_j = (X_i' - X_j')\beta + (t_i - t_j)\delta$$

In practise, the analysis will be done by the following steps. First, I geocode the sales data addresses into coordinates. Then I map each dwelling to the closest municipal border and filter out the dwellings that are not nearby any border. After that, I match each dwelling to the geographically closest dwelling on the other side of that mapped border. Following Harjunen et al. (2018), the matching is done between dwellings of the same dwelling type and sales year. Finally, I calculate the differentials of these dwelling pairs and run the regressions with them.

Figure 5 illustrates the logic of the spatial matching. Some dwellings may be the closest neighbour for more than one dwelling in the opposing side of the border. Such dwellings are therefore included several times in the pair-differenced records. However, all perfectly duplicated pairs are removed from the data.

Figure 5: Spatial matching logic

Figure 5 illustrates the logic of the spatial matching. Each dwelling is matched with the closest neighbour from the opposing side of the municipality border. In addition to geographical distance, the matching algorithms requires the dwellings to have the same building type (single-family home, row house, or apartment building) and to have been sold in the same year. Map source: OpenStreetMap



4.3. Estimating household tax payments

As discussed, the municipal tax rates are nominally flat, but effectively progressive. To analyse the magnitude of the tax capitalization, an estimation of the annual tax burden of a dwelling buying household is needed.

I design a simple estimation rule of the income level of the buyer household of a house. First, for single room apartments, I assume that the number of working residents is one. For any other dwelling types, I assume the number to be two. Second, I approximate the income of the household as a function of the dwelling price. The justified assumption in this function is that higher income is often invested in housing, i.e., expensive dwellings are bought by higher income households. Based on the Finnish household average gross salary (€55,000) (Statistics Finland, 2020) and the average dwelling price in the dataset (€184,000), I estimate the multiple to be 3.3x. Third, after calculating the income per resident, I calculate the actual annual amount of taxes by taking into account the standard tax deductions with each income level.

By this rule, for example a two-room apartment in Helsinki that costs €200,000 would result in an estimated income per resident to be €200,000 / 3.3x / 2 = €30,300. For this income, the annual municipal income taxes in Helsinki are €4,260, and then for the household in total €8,520.

This estimate is a rough one on purpose. It omits factors such as different price-to-income rations in different regions, the elasticity of income to housing spending, and the fact that the average income does not represent the average dwelling buyer income. However, a more sophisticated estimation method would likely lead to overfitting and endogeneity issues with the border discontinuity designs. A simple estimation works well with the spatial differencing method, as the differencing rules out the similar error terms of the dwelling pairs. What is interesting is the difference in taxes paid and capturing to some extent the relationship between income and spending on housing. For that, using a simple derivation from the dwelling price is sufficient.

Since the dwelling prices are highly skewed with very expensive outliers, I winsorize the derived unit-specific tax payment estimations at the 1st and 99th percentile. I find that this significantly reduces the standard errors in the models that employ the described estimated tax payment variable.

Earlier studies suggest that the taxation effect is more visible with the value of taxes to pay rather than the flat tax rate (Charlot et al., 2011). This approach of dynamic effective tax estimation might then provide more robust estimations of the capitalization. One benefit is that is automatically takes into account the salary inflation during the 21-year sample period, with the assumption that house prices have on average increased with relation to salaries. This method also allows analysing whether the capitalization rates are different in different dwelling price classes (and consequently income classes). As discussed earlier, I hypothesize that high-income households have higher motivation, capacity, and skills to take the tax factor into account, and thus the level of capitalization would be higher for more expensive apartments.

4.4. Present value of the tax savings

The net present value (NPV) of the tax savings equals the sum of discounted annual difference in taxes paid in municipalities *i* and *j:*

$$NPV = \sum_{t=1}^{\infty} \frac{tax_{j,t} - tax_{i,t}}{1/(1+r)^t},$$

where t = household taxes paid, and r = discount rate.

I next discuss how my empirical design yields an estimate of the *r* as implied discount rate (IDR), and how the level of the tax capitalization can be estimated by comparing the IDR with the economically valid discount rate for the tax savings.

4.4.1. Implied discount rate

I report the level of capitalization with the IDR of the tax differences versus dwelling price differences. This way the results can be easily compared across different specifications, and the interpretation can be done separately from estimating the valid discount rate for the tax savings.

In theory, the fully capitalized price difference of identical dwellings in municipalities *i* and *j* equals the net present value of the tax savings:

$$p_i - p_j = \sum_{t=1}^{\infty} \frac{tax_{j,t} - tax_{i,t}}{1/(1+r)^t},$$

where *p* = dwelling price. We can simplify the equation by assuming that the taxes remain constant and are paid until perpetuity:

$$p_i - p_j = \frac{tax_j - tax_i}{r}$$

As the discount rate is unknown and dependent on the empirical price differences, I call it the implied discount rate (IDR) and solve it for the taxes and prices.

$$IDR = \frac{tax_j - tax_i}{p_i - p_j}$$

4.4.2. Valid discount rate

To calculate the level of capitalization, the IDR can be compared to an estimation of what the discount rate for tax savings until perpetuity should be. Theoretically, tax liabilities should be treated as risk-free if they are mandatory and do not vary over time. In practice, municipalities do change the tax rates and the state may make changes in the taxation regime. Matikka (2018) discusses that while changes in municipal tax rates do occur, they are mostly small and predictable. Treating the tax payments as a low-risk liabilities is therefore justified.

One approach to determine the discount rate in the housing market context is to compare it to mortgage interest rates. The intuition is that if buying the dwelling is financed with a mortgage, the future tax savings can be used to pay off the loan, and therefore have the same discount rate.

The average mortgage interest rates in Finland between 2000 and 2021 has been 2.78%. (Bank of Finland, 2021)

In the tax capitalization literature, Høj et al. (2018) also argue that a discount rate of around 2-3 percent is a valid estimate for the tax payments. This is in line with observed discount rates in the housing markets as well as real interest rates of bank loans and deposits. (Borge & Rattsø, 2014)

I use the range of 2-3 % as a benchmark discount rate for the capitalization analysis. I assume that inflation offsets salary growth, and so this rate can be used as such to calculate the net present value of the tax savings until perpetuity. The level of tax capitalization is then the ratio of tax savings NPVs, calculated with the IDR and the benchmark discount rate.

4.5. Linear model biases

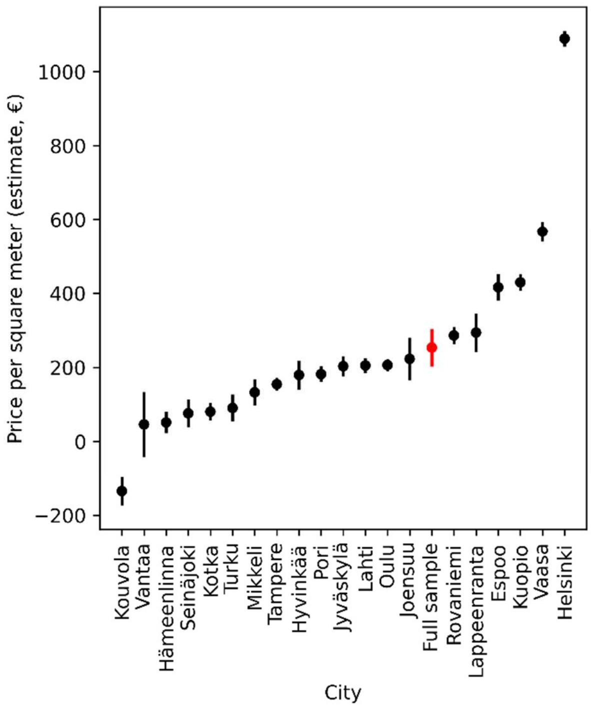
A vast majority of the econometric real estate literature and tax capitalization studies are empirically based on the linear ordinary least squares (OLS) regression. However, some of the (log-) linear assumptions in the models do not hold empirically. For example, in my dataset the *Own lot* dummy variable indicates if the dwelling is built on a freehold lot. In an OLS regression table, the coefficient for the variable represents the average price difference between dwellings on their own lots versus on rental lots. However, the sample contains data from various different locations, where the value of land is different: higher for example in Helsinki than some rural areas. This model will therefore systematically under(over-)estimate the values of dwellings on own (rental) lots in Helsinki and vice versa in rural areas. I illustrate the effect by regressing a simple hedonic pricing model separately for dwellings in the 15 largest cities in Finland, and plot the point estimates of the *Own lot* variable in Figure 6A. The results show evidently that for example in Helsinki the point estimate is significantly higher than in the full sample.

Another example of a non-linearity between dwelling features and prices is the age of the building. Generally, new dwellings are more expensive than old ones. However, in some old and central areas, dwellings from a specific era can well be the most expensive ones. On the other hand, dwellings that are approaching the need of a big renovation such as a pipe repair are clearly cheaper than dwellings that have just undergone the repair. (Nikola, 2011) I illustrate the non-linearity by plotting the price estimates for apartments sold in Helsinki between 2010-2020 in Figure 6B. The graph shows that prices do decrease rather monotonically, although not linearly, until around 50 years of age, which is the typical age for the first expensive plumbing renovation in apartment buildings. After that, the prices start to increase with relation to age, and also the variation starts to increase. As discussed in *Section 3.2.*, I try to capture this effect by using a dummy variable that indicates if a dwelling in an apartment building is over 55 years old.

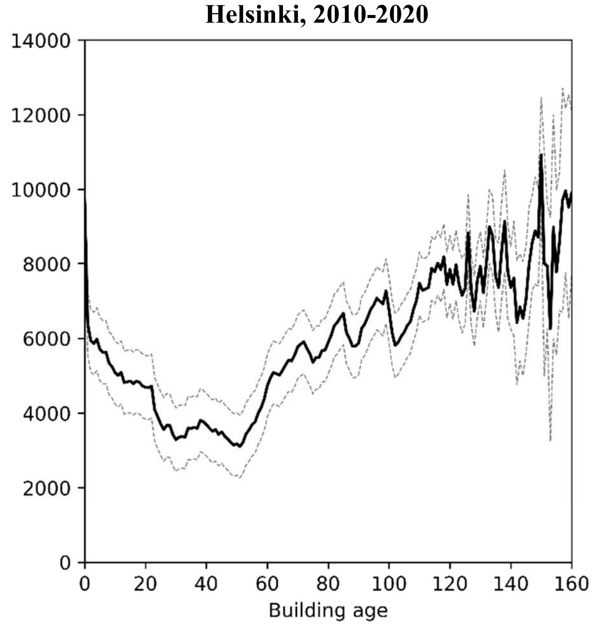
Figure 6: Dwelling pricing non-linearities

Figure 6 illustrates the non-linear relationships between lot ownership in different cities and price (Panel A); and building age and price (Panel B). Panel A specifications are constructed as subsets of the 15 largest cities in Finland. Panel B estimates come from a dynamic regression where building age is set as a categorical variable. Both figures include the point estimates and 95% confidence intervals, based on heteroscedasticity-robust standard errors.

Panel A: Lot ownership price estimates in city subsamples



Panel B: Building age price estimates in Helsinki, 2010-2020



While most of these issues can be circumvented in OLS by using a set of polynomial variables, fixed effects, or combined variables, it might not be the optimal strategy. A number of such fixes may easily overfit the model and also make the regression table impossible to interpret. In any case, the tweaks to the OLS are limited to the imagination of the researcher and may also raise questions of p-hacking.

Regression tree -based models like *gradient boosting* have been shown to significantly outperform *OLS* in house pricing prediction challenges.[1] In the next section, among other robustness checks, I describe an intuitive machine learning approach to easily solve some of the issues regarding non-linearity and variable connectedness.

4.6. Additional tests

In addition to the main model, I plan to perform a series of additional tests to analyse the existence and magnitude of the taxation effect. The tests are described in this section and the results are collected in *Section 5.2.*

4.6.2. Main model variations

Distance to border sensitivity: In my main analyses, I use a 1-kilometre threshold to identify dwellings that are located near a municipality border. As this limit is completely artificial, I also test the model with other distance thresholds between 0.25 and 2 kilometres.

Number of matched neighbours: The spatial matching baseline data is built by matching each dwelling with one closest neighbour from the opposing side of the municipal border and differencing the dwelling features. This can be altered by matching each dwelling with several neighbours and calculating the differences on the average of each feature of the matched dwellings. This method reduces the effect of chance when selecting the closest neighbour as a comparison dwelling. The trade-off is that the matching distances will increase, which might violate the assumption of identical location of the dwellings.

Dwelling price and building types: Running the model on subsets of data with determined price ranges and building types allows studying whether there are differences in the capitalization between groups of dwelling buyers. The dwelling price is a proxy for the income level or wealth of the household. The building type may reveal information of the capitalization behaviour of

---

[1] House price prediction is a popular exercise for testing machine learning models. For example, Kaggle hosts a publicly available dataset and competition of developing house pricing models (https://www.kaggle.com/c/house-prices-advanced-regression-techniques/). Some academic papers on the topic have been published as well, suggesting that machine learning models can clearly outperform OLS in house pricing accuracy. (See e.g., Truong, Nguyen, Dang, & Mei, 2020)

investors, as dwellings in apartment buildings are often rented but single-family homes and row houses are most often resided by the buyer household themselves.

Municipality and border exclusion: To analyse whether the results are driven by some specific municipality-pairs (borders) or municipalities, I construct subsamples of the data by excluding these one at a time. I limit the exclusions to those borders and municipalities that have over 1000 observations of spatially matched dwelling pairs.

4.6.1. Alternative methods

Border fixed effects: following, e.g., Black (1999) and Bayer, Ferreira, & McMillan (2007), this method is also based on a subsample of the dwellings that are located nearby municipal borders. Each dwelling is mapped to its closest border and the border (for example, "Helsinki-Vantaa") is used as a dummy variable. The regression includes the fixed effects of these borders as well as the level of taxation for each dwelling.

Machine learning pricing model: following, e.g., LeSage & Pace (2009) and Helbich, Brunauer, Vaz, & Nijkamp (2013), the impact can also be analyzed by fitting a non-linear pricing model that outputs a price prediction of each house. I use a regression-tree based *gradient boosting* model to form the predictions. The model uses the same set of control variables but does not include either of the tax variables. The model is trained using the dwellings with a distance between 1 to 2 kilometers to a municipality border and tested with the baseline dataset of distance below 1 kilometer. The *train-test split* mitigates the risk of overfitting the model to the sample. After forming the price predictions, I calculate the prediction errors (residuals) for each dwelling. The residuals are then regressed against corresponding municipal taxes. The intuition is that if capitalization exists, the predictions should have systematic errors that can be explained by the different tax rates.

Placebo data: following, e.g., Gibbons et al. (2013) and Harjunen et al. (2018), the validity of the results can be tested by creating a fake dataset with distorted municipality borders. To show that the actual methodology is valid, this specification with the fake borders should induce no effect for the tax variable. I implement this by inverting the border regions nearby the real borders and creating placebo borders at 1 kilometre distance of the real borderlines. (See, Appendix A6 for illustration). I then map the data to these new borders and run the analysis with the generated dataset.

4.7. Endogeneity concerns

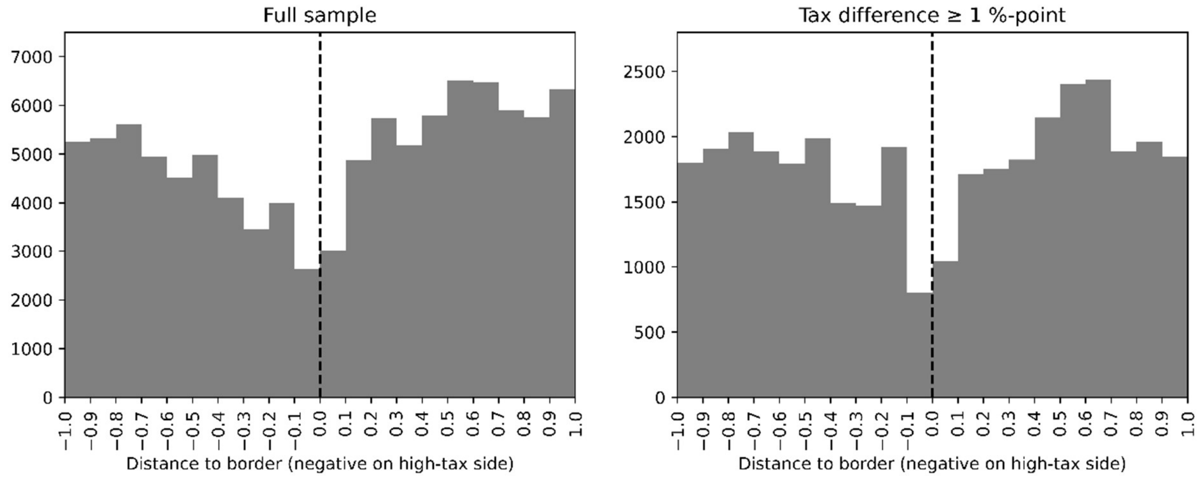
Analysing tax capitalisation between municipalities raises a fundamental endogeneity concern, as both the dwelling prices and taxes are driven by the dwelling location. However, as discussed, the border discontinuity design addresses this issue by normalizing the price impact of location at the nearby areas of the municipal borders.

The empirical design still leaves room for potential endogeneity issues. Lee & Lemieux (2010) discuss the issues specific to geographical discontinuity models. For example, it can be questioned if the district borders are in fact arbitrary or sometimes set by natural obstacles like rivers, lakes and hills or built infrastructure like streets, parks, and buildings. Furthermore, city planning is not random. In the tax capitalization setting, bigger cities tend to have lower tax rates than municipalities surrounding them. As closeness to city centre is generally a decided feature of a house, the surrounding municipalities want to locate dwellings close to the city borders, whereas the city does not have similar incentive to locate dwellings on the border regions. This would cause severe bias in the data, as the border neighbourhoods might start to differ at the different sides.

The ideal discontinuity analysis would require dwellings to pre-exist in an area, and then a set of fully arbitrary borderlines set to separate the area into pieces. This is clearly not how the borders are set in real world. Figure 7 shows that dwellings are not uniformly distributed nearby municipal borders, but instead the number of observations decrease closer to the border. However, no effect of general imbalance of dwellings located at the different sides of the border seems to exist. The distribution is approximately same at borders where the tax difference is at least 1 percentage point.

Figure 7: Number of observations nearby borders

Figure 6 illustrates the distribution of observations with relation to distance to the closest municipality border in the spatially restricted sample.



Even though the dwelling locations are not uniformly distributed, I argue that it is very unlikely that there would be such factors that would be systematic related to the tax rates. The underlying factors, like natural obstacles at borderlines, might increase in the standard errors but will not create any systematic bias that would affect the results.

Other mechanisms for the tax differences can be proposed as well. For example, municipalities might have different preferences for using different income sources, for example the income rate, property tax and service co-payments. This could make the income tax rate a noisy estimator for the total public taxes and fees that are collected within a municipality.

Another concern is that financially distressed municipalities with high tax rates might be expected to merge with their neighbouring municipality (with lower taxes). If dwelling buyers can anticipate such mean reversion in the tax rates via municipal mergers, they may not fully capitalize the tax difference on dwelling prices. Based on evidence from Finnish mergers, they do not decrease the public spending per capita but make the municipalities exploit the transformation period to pooled funds by increasing leverage and investments. (Moisio & Uusitalo, 2013; Saarimaa & Tukiainen, 2015). As the mergers cause such behaviour on the governing level, it might be visible for dwelling buyers as well. In my research setting, this effect could be further studied by restricting the data to the municipalities that have not merged during the sample period. If they have not merged, it is reasonable to assume ex-post that the market should not have priced in any merging expectations. However, because merged municipalities are most often small and sparsely populated, I do not have enough data in relevant merged borderlines to do a credible analysis with this method.

Finally, the taxation itself causes externalities that may infer the results. The relevant fiscal externality in my research context is that higher taxes cause less incentives to work by some level of elasticity. (Hendren, 2016). Matikka (2018) studies Finnish municipal income tax rates and estimates the *elasticity of taxable income* (ETI) to be 0.21. The estimate implies a relatively low sensitivity of income earners to changes in tax rates.

A more serious concern is that the effect may distort the assumed randomness in the choice of residence in the border areas: high-income households may have higher preference for the low-tax side of the border, and vice versa. While differences in the dwelling features can be controlled, I do not have access to highly granular data of household income. I estimate the buyer household's income from the dwelling prices within each border district, which means that I am bound to the assumption that income levels the same on both sides of the border. I aim to mitigate the issue with the differencing strategy. Even if the income levels differ, the differencing makes the errors significantly smaller.

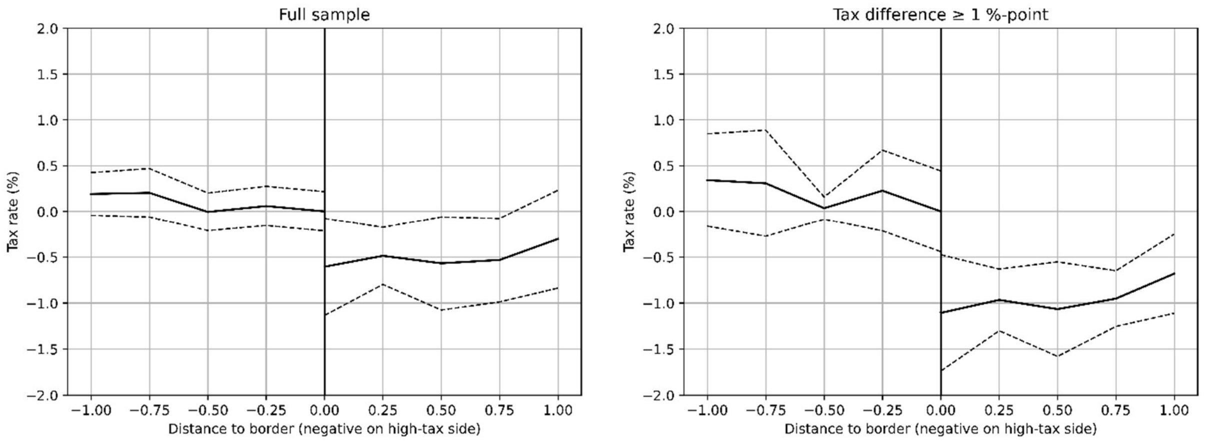
5. Results

5.1 Main results

I first provide visual evidence of how the sample dwellings close to municipal borders are identified and how the tax differences can be observed. The results are presented in *Figure 8*. The vertical centre line represents the municipality borders, and the data is divided by checking if the dwelling is located at the high-tax or low-tax side of the border. The solid line represents the average tax rate of the dwellings in each distance bin, surrounded by the 95% confidence interval. As should be due to the constructed high/low tax split, the tax differences are easily visible in the full sample figure. When the sample is restricted to boundaries with ≥ 1% tax difference, the "jump" at the boundary is even higher.

Figure 8: Border discontinuity of municipal income tax rates

Figure 8 shows the boundary discontinuity of the tax rates in the sample restricted to dwellings close to municipal borders. The solid line connects the point estimates and dashed line the 95% confidence interval of the average tax rate for each distance to the closest municipal border. Confidence intervals are based on standard errors clustered at the municipal boundary level.



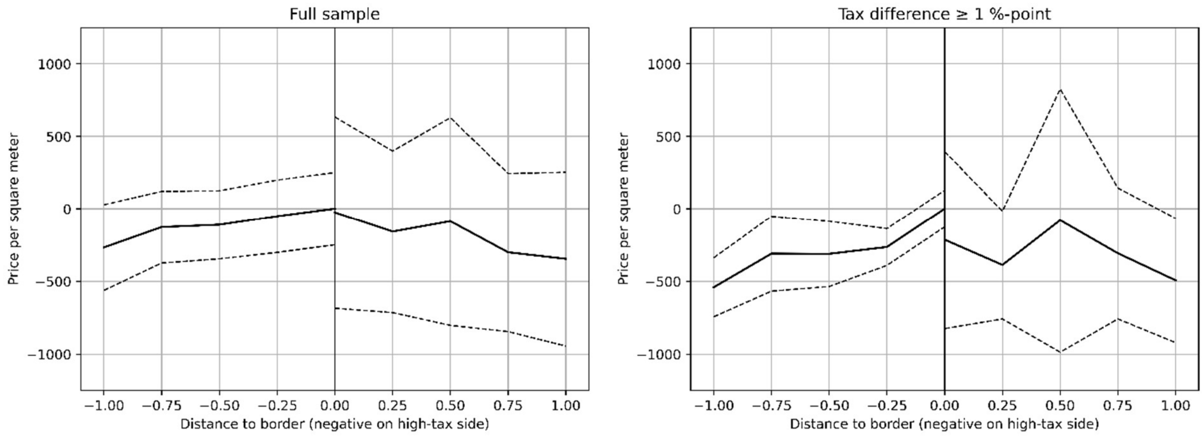
In Figure 9, I use the same construction and raw data to report the average dwelling prices (per square meter) at the opposing sides of the borders. In the uncontrolled regression (Panel A) full sample, there is basically no difference in the average prices at the opposing sides of the border. When the sample is restricted to the borders where the cross-border tax difference is greater than 1 %-point, the average prices vary more. At the border, the dwellings on the low-tax side are even slightly cheaper, contrary to the capitalization expectation. However, the difference is statistically highly insignificant.

The assumption behind comparing simple average prices is that dwellings are identical across borders. As this is likely not a valid assumption, and the dwellings might even have systematic differences across borders, I develop the model by controlling the prices with individual dwelling features in Panel B. The control variables make the standard errors slightly smaller, but results remain the same and no evidence of price difference at borders can be found.

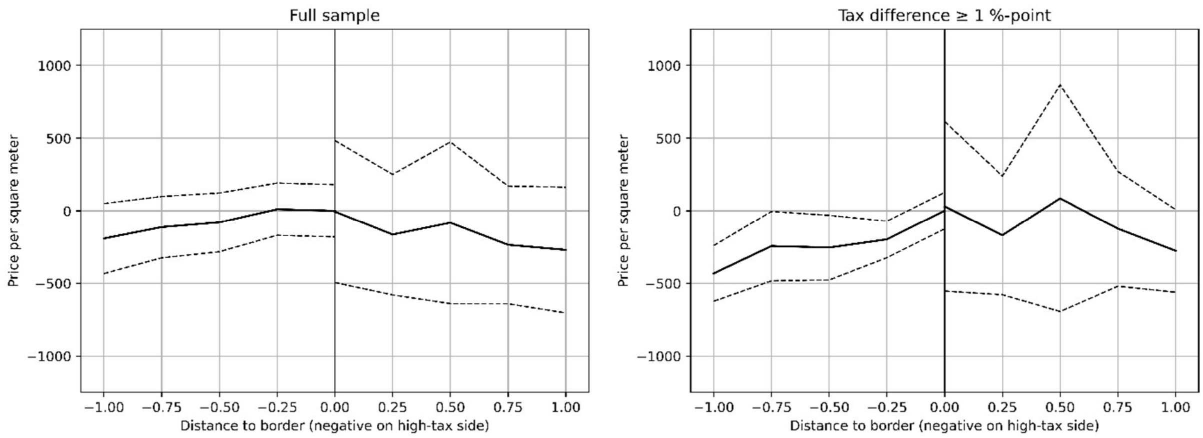
Figure 9: Border discontinuity of dwelling prices

Figure 9 shows the boundary discontinuity of the dwelling prices in the sample restricted to dwellings close to municipal borders. The solid line connects the point estimates and the dashed line the 95% confidence interval of the average prices for each distance to the closest municipal border. Confidence intervals are based on standard errors clustered at municipality boundaries.

Panel A: No control variables



Panel B: Dwelling control variables



The graphs are meant to provide simple visual evidence of the potential discontinuity, based first on the raw data and then a slightly more sophisticated version including dwelling features.

However, they are limited to controlling the features of individual dwellings and the price sensitivity to the distance to border. To take into account the most critical variable in the pricing function – dwelling location – I next present the main analysis results of the spatial matching specification.

Table 2. Spatially matched cross-boundary differences

Table 2 reports the results for the spatially differenced models. The sample consists of the dwellings that are located within 1 kilometre distance of a municipal border. *Tax estimate (€1000 per annum)* variable is winsorized at 1% and 99%. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively, based on standard errors clustered at municipality boundaries.

	Dependent variable:			
	Price / m2			
	(1)	(2)	(3)	(4)
Tax rate (%)	83.636		-66.915	
	(98.194)		(51.919)	
Tax estimate (€1000 per annum)		113.106		-193.226
		(221.392)		(135.476)
Floor area (log)			-677.640***	-683.933***
			(110.982)	(107.167)
Floor			30.285**	31.242**
			(14.539)	(14.531)
Condition: new			795.220***	786.171***
			(65.678)	(62.386)
Condition: good			477.053***	477.998***
			(42.255)	(41.386)
Condition: satisfactory			238.787***	239.374***
			(39.070)	(37.958)
Own lot			91.976	94.970
			(78.083)	(78.100)
Age			-17.824***	-17.988***
			(1.632)	(1.628)
Apartment x Age > 55			432.638***	436.453***
			(140.045)	(140.617)
Observations	99,651	99,651	99,651	99,651
Adjusted R^2	0.004	0.002	0.467	0.469

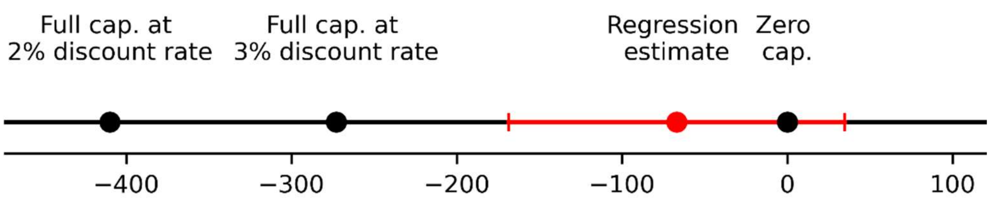
Table 2 contains the regression results for the spatially matched and differenced dwelling pairs. In the models (1) and (2) without any dwelling control variables, the tax variables show positive but insignificant estimates. However, after adding control variables, the tax variables have the expected negative coefficients, implying that higher taxes do result in lower dwelling prices. The estimates in these models (3) and (4) are still statistically not significant, so zero capitalization can not be rejected.

The interpretation of the economic significance of the estimates requires some assumptions. With model (3) plain tax rate variable, I use the average household income calculation from *Section 1.3.* to convert 1 %-unit to an annual value of €656. This is comparable to the tax value of €1000 used in model (4). The price per square meter estimates are converted into full price estimates by multiplying with the average dwelling size of 80 square meters. Finally, as reasoned in *Section 4.4.2*, I use discount rate range of 2% to 3% to calculate the present value of the tax savings and compare it to the price discount of the dwellings.

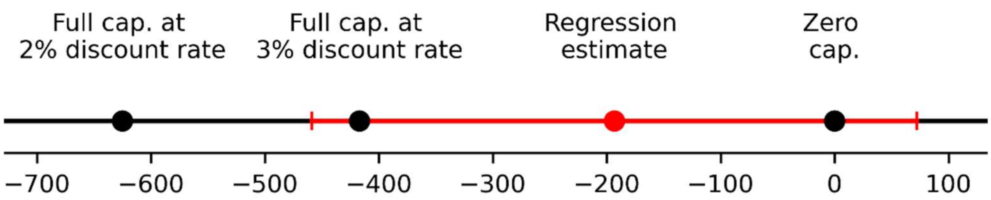
Figure 10: Main results interpretation

Figure 10 illustrates the results of the spatial matching specification. Panel A reports the results with the *Tax (%)* variable and Panel B with the *Tax (k€ per annum)* variable. The red points represent the regression point estimates for the tax variables, surrounded by the 95% confidence intervals for the estimates. Standard errors are clustered at municipal boundaries. The black points represent the cut-offs for different rates of capitalization: full capitalization with 2% or 3% discount rate, or zero capitalization.

Panel A: Model (3), *Tax rate (%)*



Panel B: Model (4), *Tax estimate (€1000 per annum)*



Using this logic, I visualize the regression results in *Figure 9*. The red dot shows the estimate and 95% confidence interval for the price per square meter -change by one percentage unit or €1000 increase in taxes. As discussed, zero capitalization can not be rejected in either of the tax variable specifications. Looking at the other tail, also full capitalization with a 2% discount rate can be rejected in both models. However, only the model (3) rejects full capitalization with a 3% discount rate as well. With the main model (4), the 3% limit is within the confidence interval and full capitalization can not be rejected.

In terms of implied discount rates (IDR), the corresponding point estimates would be 12.5% and 6.5% for models (3) and (4), respectively.

The results imply that taxes are to some extent undercapitalized. However, the confidence intervals are relatively high due to clustering standard errors at municipal boundaries.

5.2 Robustness checks

5.2.1. Distance to border sensitivity

Table 3 reports the specifications with different thresholds for the dwelling distance to a municipality border. The results are robust regardless of the distance limit. The point estimates suggest slightly higher capitalization on the 2.0, 1.5, and 0.5 km thresholds, compared to the baseline threshold of 1.0 km. However, the standard error of the estimate seems to be minimized with the 1.0 km threshold.

Table 3: Spatial matching, distance to border sensitivity

Table 3 reports the distance-to-border threshold sensitivity of the spatial matching design. All models include the dwelling control variables. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively, based on standard errors clustered at municipality boundaries.

	Dependent variable: Price / m^2				
	< 2.0 km	< 1.5 km	< 1.0 km	< 0.5 km	< 0.25 km
Tax estimate (€1000 per annum)	-288.526* (167.332)	-222.783 (148.095)	-193.226 (135.476)	-240.741 (160.788)	-188.941 (170.813)
Observations	228,433	159,769	99,651	43,477	18,301
Adjusted R^2	0.415	0.439	0.469	0.510	0.519

5.2.2. Number of matched neighbours

Next, I report the specifications with varying number of matched neighbours in Table 4. The results show that the magnitude of the tax effect seems to get a bit larger when the matching is done with more than one neighbour. With K=3, zero capitalization can be rejected at 95% confidence.

The results could be explained by the matching distances getting longer when more neighbours are added. As low-tax municipalities are typically cities with high dwelling prices, spanning the distance further away from the high-tax borderline often makes the matched apartments closer to the expensive city centre. This would be seen as a higher level of capitalization, even though the underlying reason is violating the assumption of similar location *z* of the matched dwellings. Similar effect is visible in the previous section results, where the level of capitalization increases when the distance to border threshold is increased. This hypothesis is also supported by the finding that when tested on densely populated borderlines, increasing the *K* does not have an effect in the tax coefficient.

Table 4: Spatial matching, number of matched neighbours

Table 4 reports the spatial matching design results with different number of matched neighbours. The *K* indicates how many neighbours have been matched for each house. The neighbour features have been calculated as the average of the features of the matched dwellings. All models include the dwelling control variables. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively, based on standard errors clustered at municipality boundaries.

	Dependent variable: Price / m^2		
	K = 1	K = 2	K = 3
Tax estimate	-193.226	-232.535*	-266.864**
(€1000 per annum)	(135.476)	(130.372)	(133.343)
Observations	99,651	99,651	99,651
Adjusted R^2	0.469	0.439	0.489

5.2.3. Subsets of building types and price classes

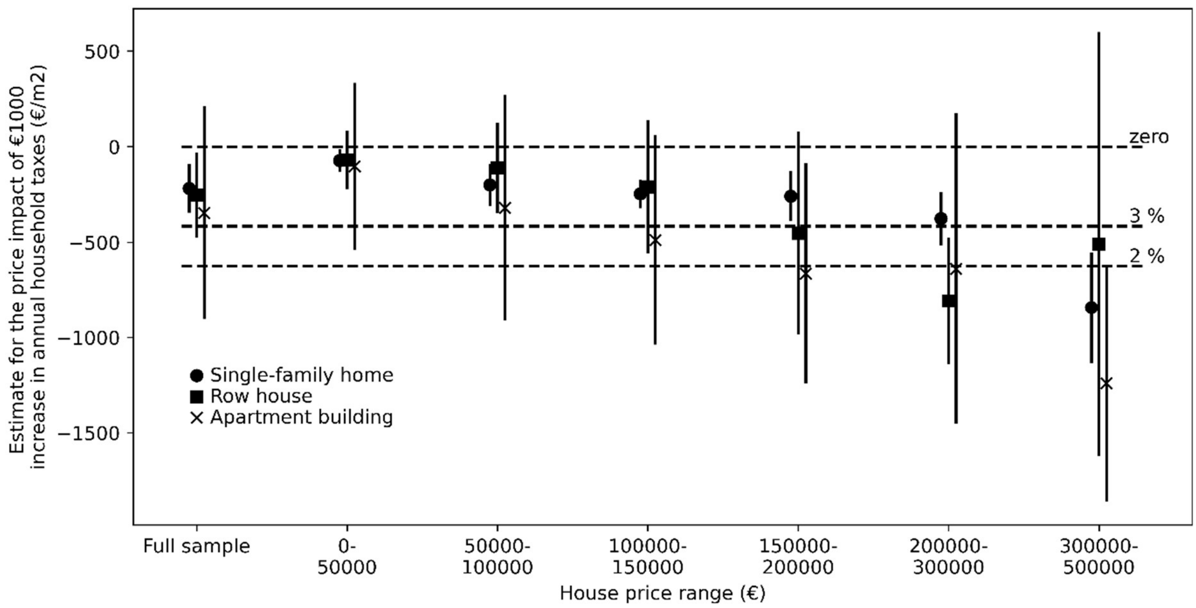
I report the specification results of subsets of different dwelling price ranges and building types in Figure 11. The interesting result is that the capitalization effect seems to be greater for more expensive dwellings. Even though the confidence intervals get wider, there is a clear downward sloping trend in the tax estimate between the price classes.

This finding suggests that more wealthy households, who assumedly buy the more expensive dwellings, have better motivation, skills, or other economical ability to take the tax factor into account when buying dwellings.

Another observation is that dwellings in apartment buildings are systematically more capitalized than single-family homes or row houses. Although the differences are not significant, one explanation could be that as apartments attract a lot of real estate investors, the average sophistication of the buyers is higher than in houses bought for own use.

Figure 11: Capitalization on different dwelling prices and building types

Figure 11 reports the spatial matching estimates of the *tax (k€ per annum)* variable on *price/m²* for subsets of data by dwelling price ranges and building types. The markers show the point estimates and the lines the 95% confidence interval of the coefficients, based on standard errors clustered at municipality boundaries. The horizontal dashed lines represent zero capitalization, and full capitalization with 3% or 2% discount rate, respectively.

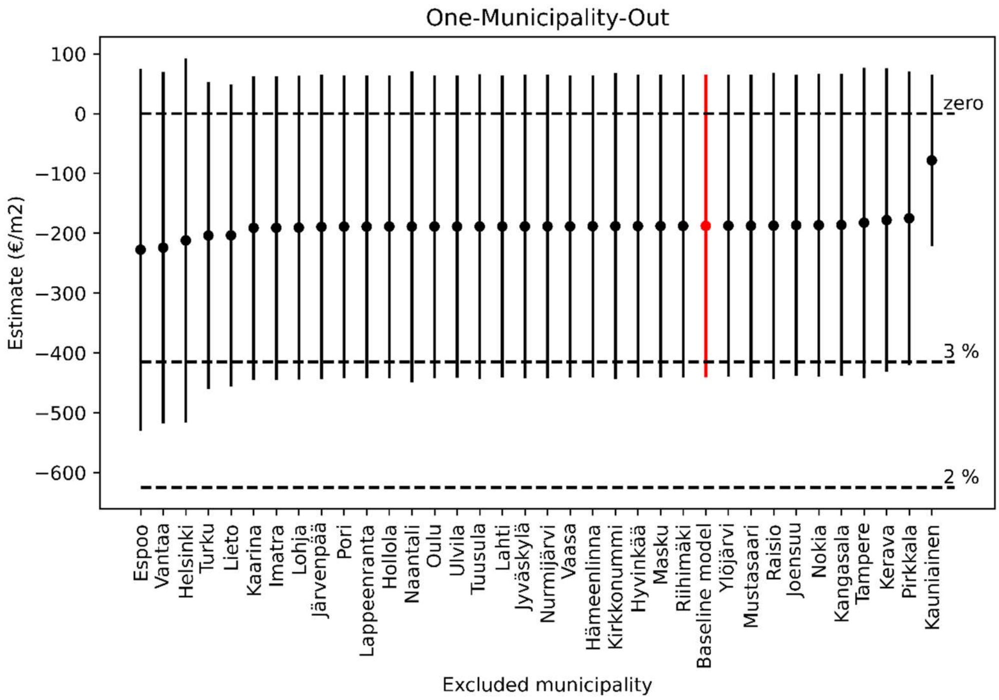


5.2.4. Subsets of municipality and border exclusions

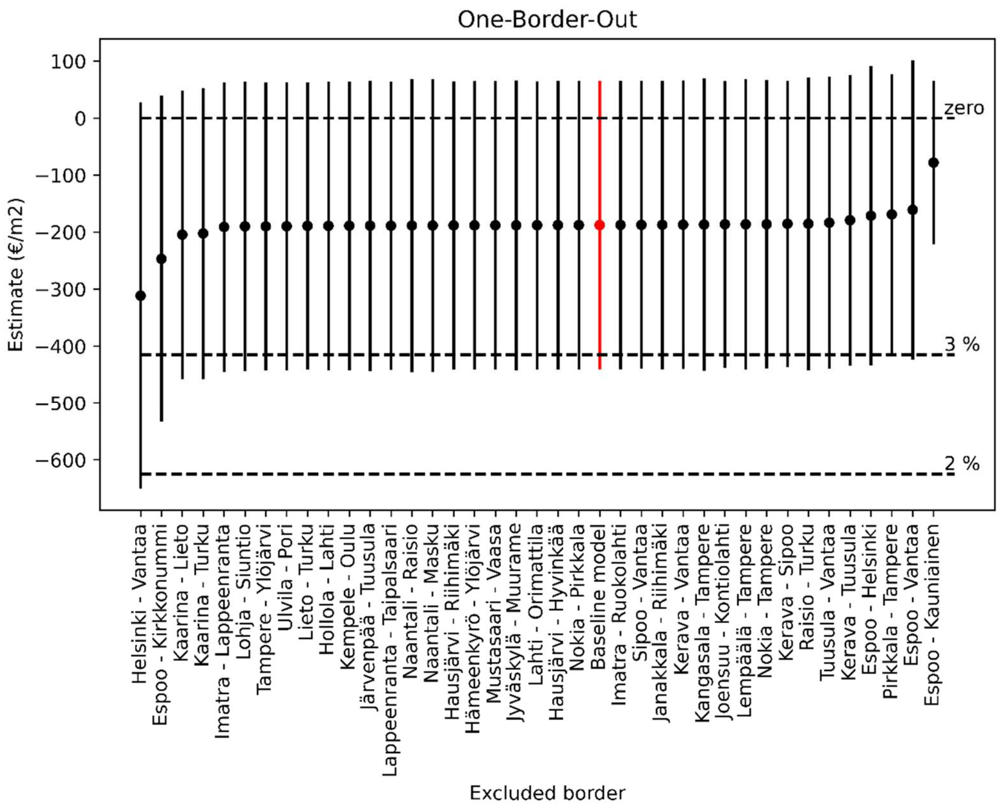
I report the specification curves in Figure 12. The results are fairly robust across the subsamples, with a notable exception that excluding the municipality of Kauniainen[1] makes the capitalization estimate significantly lower (closer to zero), makes the confidence interval narrower, and allows rejecting both the 2% and 3% discount rate thresholds for full capitalization.

Figure 12: One-Municipality-Out and One-Border-Out

Figure 12 reports the specification curve for One-Municipality-Out and One-Border-Out subsamples. The estimates come from the spatial matching specification, but in each of the subsamples excluding one municipality (or border) with over 1000 observations. The graphs include the point estimates as well as the 95% confidence interval based on standard errors clustered at municipality boundaries. The red marker highlights the baseline model with the full dataset.



[1] Kauniainen is somewhat a special case among Finnish municipalities, as it is a small town within the Espoo municipality. It is known for attracting high-income households. The average household gross income is €136k annually, 70% higher than in its only neighbour Espoo (€79k) and more than double of the average in Finland (€55k). Kauniainen also has the lowest of all municipal tax rates in Finland, 17.0% (2021).



5.2.5. Border fixed effects

In Table 4, I report the results of the border fixed effects regressions. The plain tax rate specification is in line with the spatial matching results, suggesting close to zero capitalization. However, contrary to previous results, the annual tax estimation specification has a positive coefficient. The estimate suggests reversed capitalization: higher taxes resulting in higher house prices.

The border fixed effects specification results have been tested with same robustness checks as the spatial matching method. The results are robust for subsets of data with different building types, dwelling price classes and border exclusions.

Table 5: Border fixed effects

Table 4 reports the regression summary table for the border fixed effects specification. The sample consists of the dwellings that are located within 1 kilometre distance of a municipal border. All models include the dwelling control variables. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively, based on clustered standard errors.

Dependent variable	Price / m^2	
Independent variable	Tax rate (%)	Tax estimate (€1000 per annum)
	-9.734	78.413^{***}
	(52.393)	(4.741)
Fixed effects	*closest border, building type, sales year*	
Observations	100,384	
Adjusted R^2	0.866	0.757

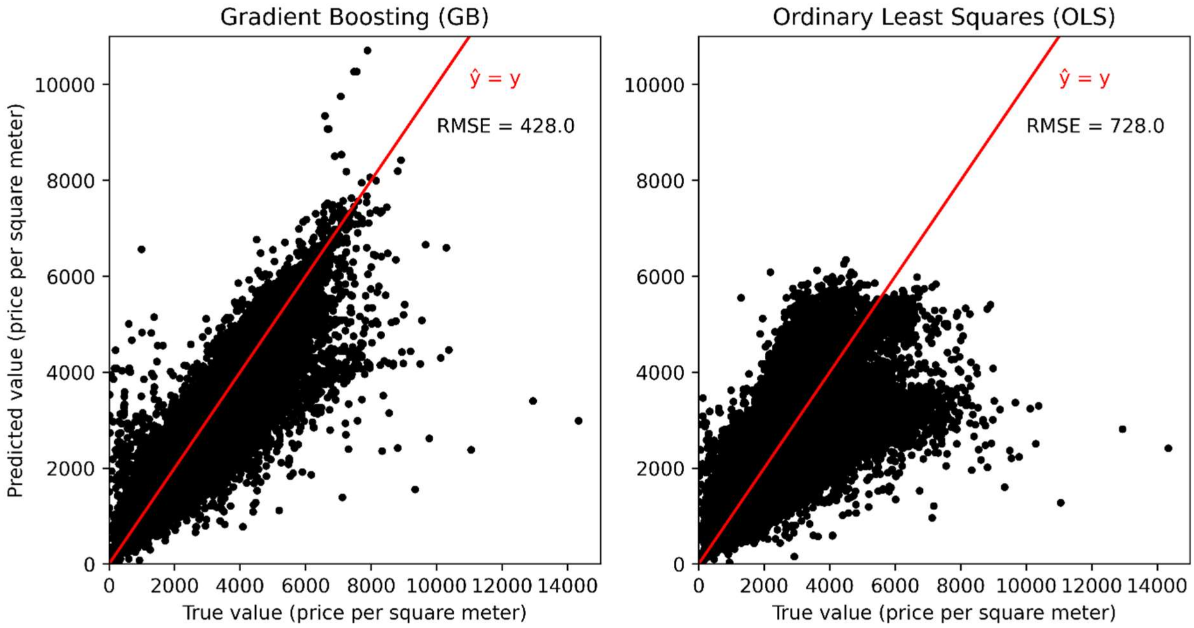
5.2.6. Gradient boosting

To rationalize the usefulness of machine learning models in the dwelling pricing domain, I first report the comparison of the gradient boosting (GB) and ordinary least squares (OLS) models in dwelling pricing accuracy. The models are trained using the distance range 1 to 2 kilometre and tested with the range of 0 to 1 kilometre from the closest municipality border. Both models include the same set of control variables. The results in Figure 13 show that the gradient boosting model clearly outperforms OLS in out-of-sample accuracy, when measured by root mean square error (RMSE).

In the specifications reported earlier, the standard errors are relatively high and make the interpretation of the results challenging. A motivation for using a more precise pricing algorithm is therefore clear.

Figure 13: Gradient boosting and OLS prediction accuracy

Figure 13 reports the predicted values plotted against the true values (prices per square meter), and the RMSE scores for the gradient boosting (GB) and OLS models. Both models are trained using the distance range 1-2 km and tested with the range of 0-1km from the closest municipality border. The GB model is implemented with *CatBoost* package and default model parameters. The OLS model is trained with a log-transformed dependent variable. The red diagonal indicates the perfect prediction line ($\hat{y} = y$).

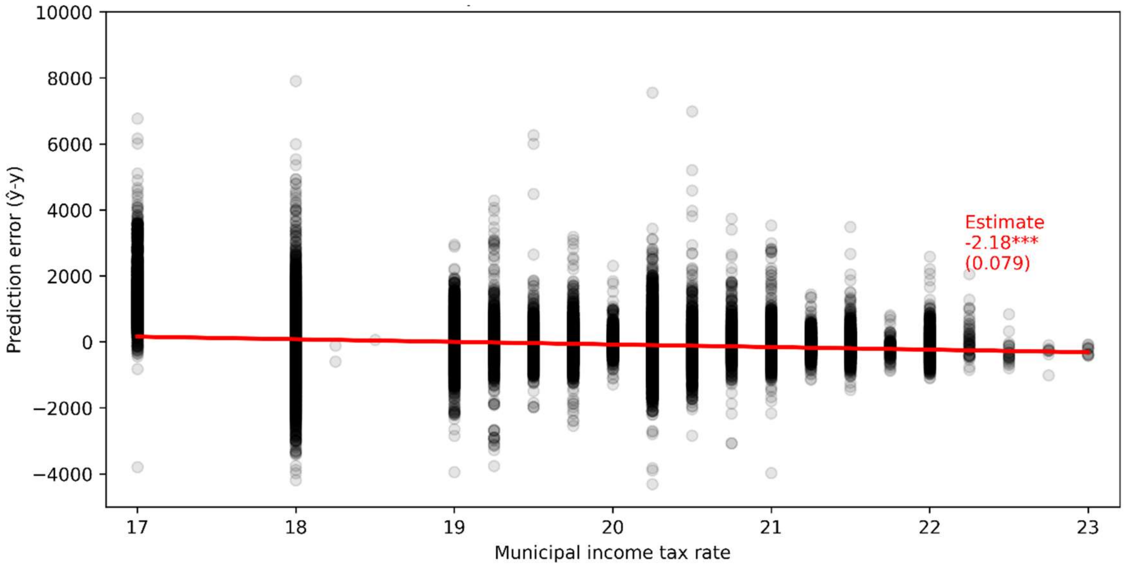


To analyse the capitalization, I regress the prediction residuals from the GB models against the tax rates. The intuition is that as the GB model is trained using the municipality borders as categorical variables, the model does not contain any information about the tax rate differences at the different sides of the borders. Any tax effect should then be included in the model prediction residuals. By regressing the residuals against tax rates, I can capture the remaining tax effect.

I present the results in Figure 14. In most of the discrete tax rate bins, the prediction errors are systematically centered around zero. The point estimate for the tax rate is -2.18 euro decrease in price per square meter for 1 %-point increase in the tax rate. The results therefore clearly suggest close to zero capitalization. The effect of Kauniainen is again visible, as it is the only municipality with a tax rate of 17%. The prediction errors are clearly positive, indicating that the pricing model can not explain the high prices. Following the intuition of this design, the abnormally high prices can be explained by the low tax rate.

Figure 14: Gradient boosting prediction residuals and tax rates

Figure 14 reports the GB model out-of-sample residuals (prediction errors) plotted against the municipal income tax rates. The red line indicates the linear fit between the variables.



5.2.7. Placebo data

Finally, I report the results of the placebo data specification of the spatial matching design in Table 5. As expected, the coefficients of the tax variables are statistically insignificant. Compared to the baseline model, the estimates are also much closer to zero, and both 2% and 3% capitalization rates can be rejected with high confidence.

Table 6: Spatial matching with placebo data

The table reports the results for the spatial differencing model using a placebo dataset. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively, based on standard errors clustered at municipality boundaries.

Dependent variable	Price / m^2	Price / m^2
Independent variable	Tax (%)	Tax (€1000 per annum)
	-29.060	-13.377
	(28.335)	(64.259)
Observations	127,795	127,795
Adjusted R^2	0.467	0.469

The placebo data results are reassuring for the validity of the other results reported in this paper. Despite the high standard errors in the spatial matching specifications, the true tax differences systematically induce clear difference in house prices, whereas the placebo differences do not.

6. Discussion

In this paper, I use various methodological approaches to study the level of municipal income tax capitalization on dwelling prices in Finland. While neither zero nor full capitalization can be rejected with high robustness, the results consistently suggest that municipal income taxes are to some extent undercapitalized on housing prices. As a simplistic rule of thumb, I estimate that the capitalization is about half of the valid present value of the taxes.

Although the main findings are reproduced on several different specifications, I also find that there is significant variation in the level of capitalization between geographical regions and house price classes. For example, the famously wealthy municipality of Kauniainen is dragging the results towards high capitalization, and without it, the point estimate of the level of capitalization would be only around 25 percent. Similarly – and highly correlated with the previous note – I find that the level of capitalization seems to increase with relation to the house price. The intuitive explanation is that wealthy households have higher motivation and better skills to take taxation into account when buying housing.

The results have interesting practical implications. From an individual point of view, a house buyer should prefer low-tax municipalities to exploit the undercapitalization. Going back to the case example in *Section 1.3* and using the discovered level of capitalization, the tax savings by selecting a dwelling at the low-tax side of the border would be around €10,000. This is a significant pricing error for most individual households, relative to their typical annual net income. It is also worth to emphasize that although the empirical design of this study utilized the neighbourhoods nearby municipal borders, the conclusions of mispricing are generalizable to all areas.

The implications also go beyond individual house buyers. Peoples' responsiveness to changes in tax rates is in the interest of decision makers, in this case on the municipal governing level. My results support the earlier findings that the responsiveness or elasticity rarely fully offsets the changes and may even be relatively low in Finland.

My contribution to the subject leaves interesting opportunities for further research. For example, the capitalization could be studied on apartment rents instead of selling prices. Also, municipal mergers or other jurisdictional events could be used as exogenous shocks to see how the followed change in tax rate is capitalized on dwelling prices.

Furthermore, it would be interesting to study the behavioural factors behind undercapitalization. For example, with a gamified experiment where test subjects are asked to make dwelling buying decisions with varying information, the pricing errors could be decomposed into known biases such as anchoring, attribute substitution or hyperbolic discounting.

Methodologically, real estate econometrics could benefit a lot from applying machine learning models instead of OLS regressions. My contribution in this paper provided one intuitive approach in utilizing such models, and hopefully inspires future work with more advanced methodological designs.

Acknowledgements

I would like to thank Michael Ungeheuer and Tapio Haaga for their comments and excellent advice during the thesis process. Furthermore, I thank Kiinteistönvälitysalan keskusliitto (KVKL) and Kuntaliitto for their support in providing the data on dwelling sales transactions and municipal tax rates.

References

Bank of Finland. (2021). MFI balance sheet (loans and deposits) and interest rates. Retrieved from https://www.suomenpankki.fi/en/Statistics/mfi-balance-sheet/

Bayer, P., Ferreira, F., & McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, *115*(4), 588–638.

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, *114*(2), 577–599.

Black, S. E., & Machin, S. (2011). Housing valuations of school performance. In *Handbook of the Economics of Education* (Vol. 3, pp. 485–519). Elsevier.

Borge, L. E., & Rattsø, J. (2014). Capitalization of Property Taxes in Norway. *Public Finance Review*, *42*(5), 635–661.

Charlot, S., Paty, S., & Visalli, M. (2011). Assessing the impact of local taxation on property prices: a spatial matching contribution. *Http://Dx.Doi.Org/10.1080/00036846.2011.613798*, *45*(9), 1151–1166.

Cushing, B. J. (1984). Capitalization of interjurisdictional fiscal differentials: An alternative approach. *Journal of Urban Economics*, *15*(3), 317–326.

Elster, J. (1979). Ulysses and the sirens: Studies in rationality and irrationality. *Cambridge University Press*.

Follain, J., & Malpezzi, S. (1981). The flight to the suburbs: insights gained from an analysis of central-city vs suburban housing costs. *Journal of Urban Economics*, *9*(3), 381–398.

Gallagher, R. M., Kurban, H., & Persky, J. J. (2013). Small homes, public schools, and property tax capitalization. *Regional Science and Urban Economics*, *43*(2), 422–428.

Gibbons, S., Machin, S., & Silva, O. (2013). Valuing school quality using boundary discontinuities. *Journal of Urban Economics*, *75*(1), 15–28.

Giertz, S. H., Ramezani, R., & Beron, K. J. (2021). Property tax capitalization, a case study of Dallas County. *Regional Science and Urban Economics*, *89*, 103680.

Harjunen, O., Kortelainen, M., & Saarimaa, T. (2018). Best Education Money Can Buy? Capitalization of School Quality in Finland. *CESifo Economic Studies*, *64*(2), 150–175.

Heinämäki, L. (2011). *Yhteistoiminta-alueiden sosiaali-ja terveyspalvelut 2010: Järjestämisen, tuottamisen ja hallinnon kysymyksiä uusissa palvelurakenteissa*.

Helbich, M., Brunauer, W., Vaz, E., & Nijkamp, P. (2013). Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria: *Http://Dx.Doi.Org/10.1177/0042098013492234*, *51*(2), 390–411.

Hendren, N. (2016). The Policy Elasticity. *Https://Doi-Org.Libproxy.Aalto.Fi/10.1086/685593*, *30*(1), 51–89.

Hipsman, N. E. (2018). *The Effect of State Income Taxes on Home Values: Evidence from a Border Pair Study*. Retrieved from www.jchs.harvard.edu

Hirshleifer, D. (2001). Investor Psychology and Asset Pricing. *The Journal of Finance*, *56*(4), 1533–1597.

Høj, A. K., Jørgensen, M. R., & Schou, P. (2018). Land Tax Changes and Full Capitalisation. *Fiscal Studies*, *39*(2), 365–380.

Holm, P., & Huovari, J. (2011). Kunnat verouudistuksen maksajina? *Kunnallisalan Kehittämissäätiön Tutkimusjulkaisu , PTT Raportteja*.

Kalluinen, J. (2021). *Kuntien verot 2021 - missä maksat eniten? Laskelmia kunnallisista veroista maakunnissa, niiden keskuskaupungeissa ja Helsingin seudulla*.

Kauko, T. J. (2002). *Modelling the locational determinants of house prices: neural network and value tree approaches*.

Keele, L. J., & Titiunik, R. (2015). Geographic Boundaries as Regression Discontinuities. *Political Analysis*, *23*(1), 127–155.

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, *48*(2), 281–355.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Lyytikäinen, T. (2012). Tax competition among local governments: Evidence from a property tax reform in Finland. *Journal of Public Economics*, *96*(7–8), 584–595.

Mäenpää, N., & Mäenpää, M. (2021). *The effect of maintenance charges on housing prices in Finland*.

Matikka, T. (2018). Elasticity of Taxable Income: Evidence from Changes in Municipal Income Tax Rates in Finland. *The Scandinavian Journal of Economics*, *120*(3), 943–973.

McMillan, M., & Carlson, R. (1977). The Effects of Property Taxes and Local Public Services upon Residential Property Values in Small Wisconsin Cities. *American Journal of Agricultural Economics*, *59*(1), 81–87.

Ministry of Education and Culture. (2019). PISA 2018: Finland has top readers. Retrieved from https://okm.fi/-/pisa-2018-suomi-lukutaidossa-parhaiden-joukossa?languageId=en_US

Moisio, A., & Uusitalo, R. (2013). The Impact of Municipal Mergers on Local Public Expenditures in Finland. *Public Finance and Management*. Retrieved from https://www.researchgate.net/publication/272795680_The_Impact_of_Municipal_Mergers_on_Local_Public_Expenditures_in_Finland

Nikola, N. (2011). *The effect of pipe repairs on housing prices*.

Oates, W. E. (1969). The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the Tiebout hypothesis. *Journal of Political Economy*, *77*(6), 957–971.

Palmon, O., & Smith, B. A. (1998). New Evidence on Property Tax Capitalization. *Journal of Political Economy*, *106*(5), 1099–1111.

Peni, L., Riipinen, T., Tukiainen, J., & Viitala, T. (2021). Kuntien verojärjestelmä: kannustimet, kuntien väliset erot ja skenaario verotuottojen kehityksestä. *Valtioneuvoston Selvitys- Ja*

Tutkimustoiminnan Julkaisusarja, (2021:9).

Rantanen, H. (2019). *The role of housing company loans in apartment price determination – Empirical evidence from largest cities in Finland*. Retrieved from https://aaltodoc.aalto.fi:443/handle/123456789/43490

Reinhard, R. M. (1981). Estimating property tax capitalization: A further comment. *Journal of Political Economy*, *89*(6), 1251–1260.

Saarimaa, T., & Tukiainen, J. (2015). Common pool problems in voluntary municipal mergers. *European Journal of Political Economy*, *38*, 140–152.

Salonen, V., & Nissinen, M. (2019). *Verotuloihin perustuva valtionosuuden tasaus – mitä tasataan ja miten?* Retrieved from https://valtioneuvosto.fi/-/10623/verotuloihin-perustuva-valtionosuuden-tasaus-mita-tasataan-ja-miten-

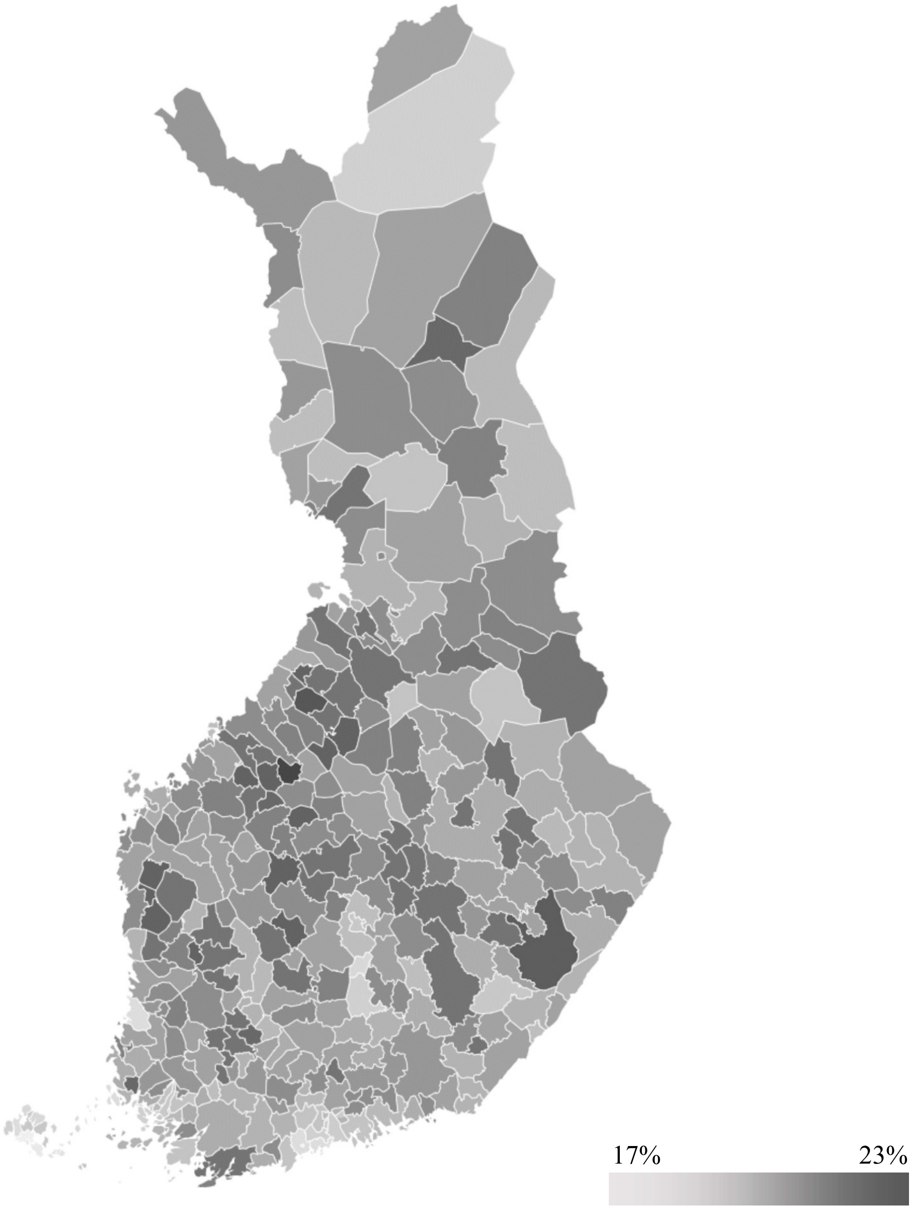
Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, *64*(5), 416–424. Retrieved from http://www.jstor.org/stable/1826343

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, *174*, 433–442.

Viherkenttä, T. (2021). *Kuntien verotulorahoituksen näkymät ja kehittämistarpeet*. Retrieved from http://urn.fi/URN:ISBN:978-952-367-490-5

Appendix

Appendix A1: Finnish municipal tax rates on map



Source: Association of Finnish Taxpayers

Appendix A2: Examples of municipal borders splitting neighbourhoods

Helsinki – Espoo

Tampere – Kangasala

Source: OpenStreetMap (© OpenStreetMap contributors)

Appendix A3: Analysis of tax rate convergence

My baseline design of calculating the present value of tax savings is based on the assumption of the tax rate difference between municipalities staying constant until perpetuity. This assumption is based on the difficult predictability of municipal taxes, and also the heuristic approach to tax difference capitalization that a dwelling buyer would most likely have. However, if the tax rate differences would systematically increase or decrease (diverge or converge), the present value calculation would be biased.

The net present value of the tax difference is calculated as

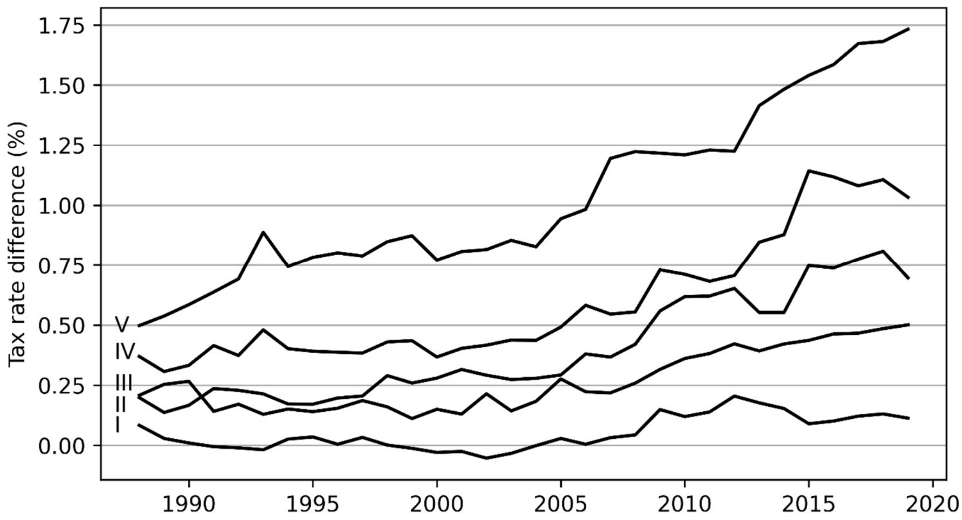
$$NPV = \frac{tax_j - tax_i}{r},$$

where r = constant discount rate. If the tax rates are not constant, but the difference grows *g* percent annually, the growth rate *g* must be added in the calculation:

$$NPV = \frac{tax_j - tax_i}{r - g}.$$

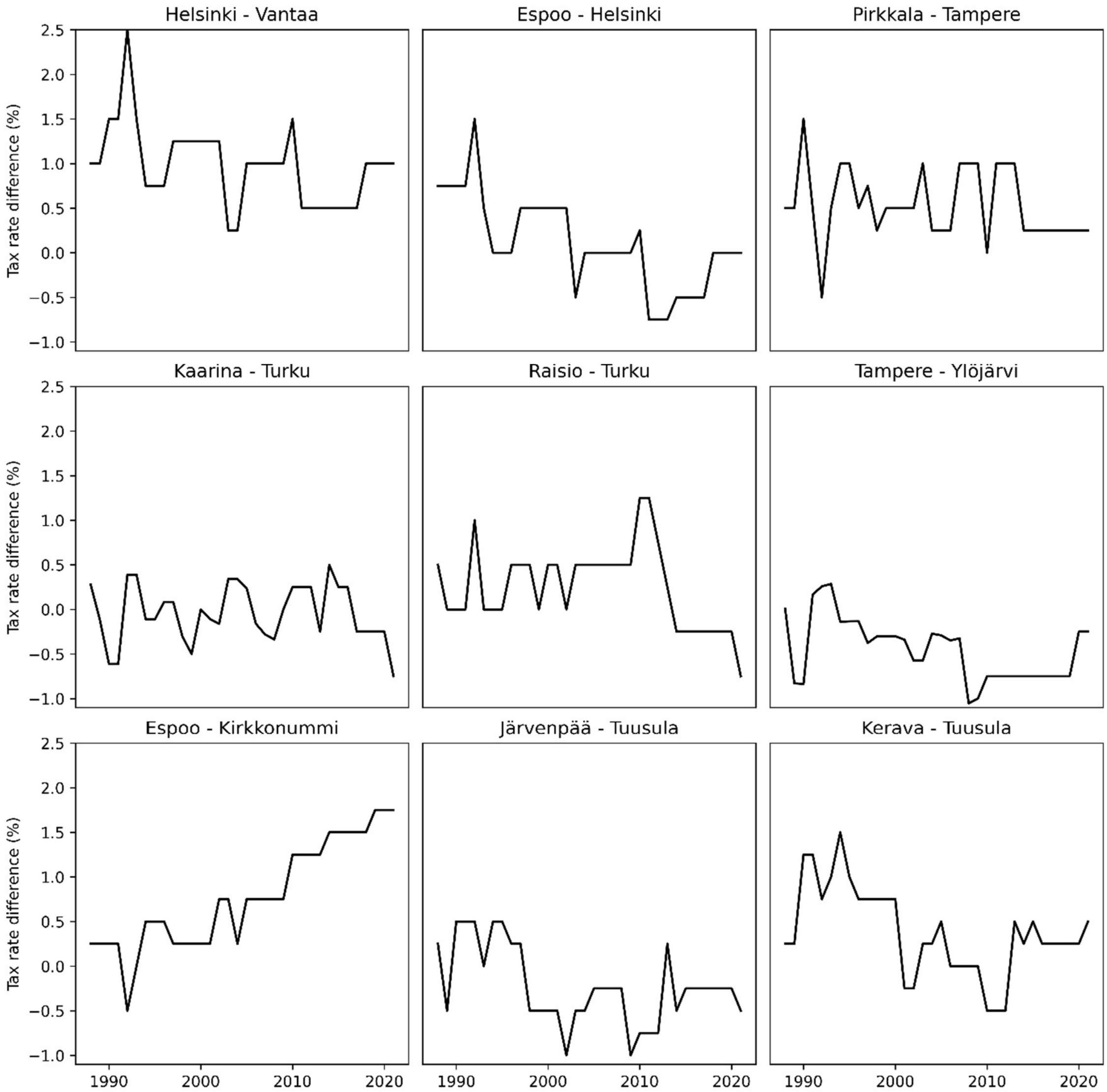
I study the tax rate convergence by dividing the pairwise differenced tax rates of neighbouring municipalities into quintiles on each year. Next, I calculate the average of the differences in each quintile for the years until 2021. I repeat this for each year between 1988 and 2019 and report the average values in the figure below.

The results indicate that overall, the tax rate differences have been growing in recent years. The highest differences in quintile V have grown from 0.5% to 1.75%, and the growth rate is lower for other quintiles.

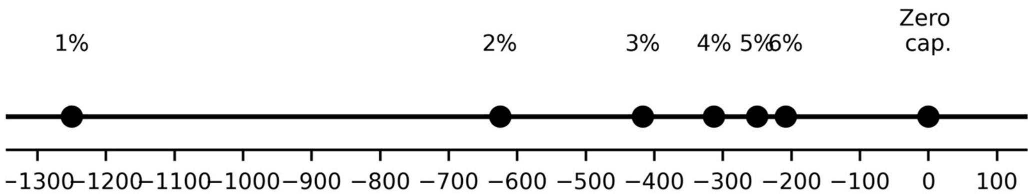


Next, I study the pairwise tax rate differences of single municipal pairs. I visualize the 9 most densely populated borderline areas, based on the frequencies in my dataset. Based on visual judgement, there seems to be quite a bit of volatility in the tax rate differences. Also, compared to the base year 1988, in 2 pairs the difference has grown, 6 decreased, and 1 stayed the same.

The results suggest that while volatility exists, the tax differences seem to be mean reverting within the municipal pair. However, some level of converge towards zero can be observed.



To illustrate the sensitivity of the main results (*tax estimate €1000*) to the tax rate difference growth rate *g*, I plot the thresholds for full capitalization at different levels of the discount factor (r-g). All in all, the observed tax convergence implies a negative growth rate *g* and therefore a discount rate of slightly above 3%. The limits for 4% to 6% rate capitalizations are close to the main specification estimates, but overall, the main findings hold even if the discount rate is increased due to tax convergence.



Appendix A4: Data cleaning process

Variable	Process	Rationale
Building type	Removed all non-residential buildings. Mapped buildings to three main categories: *apartment buildings, row houses and single-family homes*	Only residential buildings are affected by the municipal income tax rate. Reducing the number of building types allows spatial matching with a smaller radius.
Living area	Removed dwellings with area <20 or >500	Minimum dwelling area is set by law. Capping the maximum removes some excess outliers.
Floor	Single-family houses and row houses set to 1.	Floor-variable indicates the number of floors in single-family homes and row houses, which is not as relevant as the *floor number* of an apartment in a block house.
Price	Winsorized at 1% and 99%	Clearing outliers in both the price variable and the tax value estimation.

Appendix A5: Descriptive statistics

Distance to a municipality border < 2 km
N = 231,502

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
living_area	80.544	191.050	14	55.5	96	90,000
building_year	1,983.707	20.193	1,657	1,971	2,001	2,022
sales_year	2,012.065	5.451	2,000	2,007	2,017	2,021
age	28.358	20.024	-1	12	41	347
number_of_rooms	2.993	1.336	1.000	2.000	4.000	100.000
floor	1.926	1.543	1	1	2	23
own_lot	0.754	0.431	0	1	1	1
price	171,705	113,021	47	100,000	216,600	8,400,000
price_per_sqm	2,243	1,096	0.810	1,476.190	2,783.020	43,678.160
tax_rate (2021)	19.632	1.100	17.000	19.000	20.500	23.500

Appendix A6: Placebo municipal borders construction

The figure below illustrates the logic of forming the placebo municipal borders. The solid black line represents the true borderline between two municipalities (tax areas). In the right-hand-side figure, the placebo borders are marked with dashed red lines. The borders are constructed by inverting the areas within 1 kilometre distance of the true border. The spatial matching in the placebo specification is done across the placebo borderlines, not the true borders.

True border **Placebo borders**