Statistisk sentralbyrå
Statistics Norway

Hedonic regression models for housing tax valuation

Erlend Eide Bø, Odd Erik Nygård & Thor Olav Thoresen

TALL SOM FORTELLER

DISCUSSION PAPERS

1019

Discussion Papers: comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

The Discussion Papers series presents results from ongoing research projects and other research and analysis by SSB staff. The views and conclusions in this document are those of the authors.

Published: December 2024
Corrected 20 December 2024: Figure 2, page 16

Abstracts with downloadable Discussion Papers in PDF are available on the Internet:

https://www.ssb.no/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Abstract

Different types of taxation include the market value of housing or housing returns in their tax base, making it essential to obtain accurate and up-to-date assessments of property values. However, to value residential property represents a major challenge for tax administrations due to informational constraints. In the present paper we present and discuss a simple, inexpensive, and transparent procedure for assigning market value to each dwelling in Norway, based on deriving estimates from hedonic regressions. The valuations are updated yearly to reflect changes in market value. This is a novel example of using predictions obtained from regression estimates to define full-scale housing values for tax purpose. We present and discuss two iterations of the method: the initial prediction model introduced in 2010 and a refined version that would offer substantial improvements without the need for additional data collection efforts.

Keywords: Taxation of housing, Hedonic regression, Housing valuation

JEL classification: C51, D31, H61

Acknowledgements: We thank Pascal Zamorski and seminar participants at presentations at the Ministry of Finance (April, 2024) and the 2024 IIPF conference for comments to a previous version of the paper. Financial support from the Ministry of Finance is gratefully acknowledged.

Address:

Erlend Eide Bø, Statistics Norway, Research Department. E-mail: erlend.bo@ssb.no

Odd Erik Nygård, Statistics Norway, Research Department. E-mail: odd.erik.nygard@ssb.no

Thor Olav Thoresen, Statistics Norway, Research Department. E-mail: thor.olav.thoresen@ssb.no

Sammendrag

Å ha et landsomfattende system for verdsettelse av boliger er krevende, men er viktig blant annet fordi boliger danner grunnlag for beskatning. Systemet bør helst gi korrekte og oppdaterte prediksjoner for boligers markedsverdi. I Norge har boligverdier vært brukt i flere typer beskatning: skatt på fordel av egen bolig, arveavgift, formuesskatt og eiendomsskatt. Skatt på fordel av egen bolig ble avskaffet i 2005 og arveavgiften ble borte i 2014. I mange land brukes store ressurser på boligverdsetting, mens andre land bruker gamle, utdaterte boligverdier.

Hovedbudskapet i denne artikkelen er at et landsomfattende, relativt presist system for verdsettelser av boliger ikke behøver å være verken komplisert eller dyrt. Det norske systemet for boligverdsettelser, som ble etablert i 2010, er et eksempel på opprettelse av et system som både er enkelt å administrere og billig å drive.

Kjernen i dette systemet er en hedonisk boligprismodell, utviklet i SSB. Informasjon om boligpriser og attributter fra omsatte boliger blir hentet fra FINN.no, og brukt til å estimere sammenhengen mellom boligpriser og observerbare boligkjennetegn. Denne informasjonen anvendes deretter til å predikere verdier for alle norske boliger. Verdiene formidles til skattyterne gjennom skattemeldingen.

Forut for implementeringen av dette verdsettelsessystemet i 2010, var det en erkjennelse av at det eksisterende systemet verken ga korrekte verdier eller var rettferdig. For eksempel ble dyre leiligheter på Frogner i Oslo verdsatt svært lavt på grunn av lave historiske verdier. Vi går gjennom den politiske prosessen rundt innføringen av den eksisterende modellen.

Artikkelen presenterer to forskjellige versjoner av hedoniske prismodeller. Den første modellen er den som anvendes av Skatteetaten i dag for verdsettelse av boliger. Den andre modellen er en forbedret utgave, basert på finere geografisk oppdeling av datamaterialet. Den siste varianten øker treffsikkerheten betraktelig, uten å kreve mer tilgjengelige data. Treffsikkerheten øker spesielt for verdsettelse av de dyreste boligene, som til nå har vært til dels kraftig undervurdert.

1 Introduction

There are several potential types of taxation that rely on housing values, such as the recurrent tax on property, tax on imputed housing rent as part of the personal income tax, annual taxation of recurrent wealth, and taxes on intergenerational transfers (gifts and bequests). Numerous countries impose one or more of these taxes, and a challenge arises regarding how governments assess residential property.

Valuation of housing is a major administrative problem as the market values of a large majority of houses are typically unobservable. Consequently, numerous valuation methods rely on either outdated historical housing appraisals or relatively expensive systems involving personal assessments by inspectors (Almy, 2014). For example, in both the UK and Germany, historical values are utilized, while the prevailing method in the US is based on valuations by tax administration inspectors.

The main message of this study is that gathering information on housing values can be straightforward and affordable. The housing valuation system of Norway illustrates how asset values can be derived using transaction data and basic regression methods, all while keeping administrative and compliance costs relatively low. The contribution of this paper lies in describing the development and implementation of this approach for a real-world application – taxation of housing wealth.

Since 2010, the Norwegian Tax Administration has calculated housing values for the entire housing stock in Norway for tax purposes using a hedonic regression model based on housing transaction data. The housing values are updated yearly with data on new transactions. In the present study we demonstrate how market values are obtained using predicted values from the hedonic regression analysis and describe how these values are used in a practical valuation system for tax purposes.

After a decade of use, two main problems with the existing version of the hedonic model has resulted in initiatives to revise the valuation model. First, because the procedure uses the municipality as a primary economic unit (i.e., model estimates are obtained for groups of municipalities), it is not robust to changes in the municipality structure.[1] Second, the present procedure has received criticism because of lack of accuracy. In particular, there are indications that very expensive houses being under-valued by the procedure (Takle and Medby, 2020).

As a result, an improved model has been developed, but not yet implemented. In the following we also present this revised model and its prediction results. A key innovation of the new model is that information is collected at a more granular level than the municipality, using information at the level of the "basic statistical unit" (grunnkrets). The model thus better incorporates price differences between different neighborhoods within municipalities. Whereas there are approximately 350 municipalities in Norway at the moment, there are about 14,000 basic statistical units. Machine learning techniques are used to group these units to improve the empirical model.

Using information from housing transaction data and econometric techniques to obtain knowledge on regularities behind values is not new. "Mass valuation systems" (or "CAMA systems") may involve the development of models with the ambition to compute each property's value (Eckert, 2008; Norregaard, 2013; Almy, 2014). An important reason for adopting a regression-based schedule in the Norwegian context was that it replaced a highly deficient valuation scheme. Before 2010

[1] Recently, a large number of municipalities were merged in Norway.

housing values were obtained from a register of historical tariff values, where housing units were established by cadastral information. As housing was often valued at construction cost, an implication was that older dwellings had unrealistically low values, with corresponding legitimacy problems of the tax schemes based on these values.

The main reason for establishing a reliable system for obtaining housing values is the functioning of the tax system. The literature highlights both equity and efficiency arguments for eliminating the preferential treatment of homeowners, which is a common phenomenon throughout the world. Reforming housing taxation can serve as a viable means for governments to increase tax revenues without negatively impacting income inequality or work incentives (Figari, Paulus, Sutherland, Tsakloglou, Verbist, and Zantomio, 2017). Correct valuation of housing is an important prerequisite for successful implementation of property tax reforms (Grover, 2016).

The taxation of housing can have a progressive effect on the redistributional effect of the tax system (Bø, 2020). Though if outdated or erroneous housing values does not reflect current market prices, it is less likely to be, as one would expect housing values to be closely correlated to income. Recent literature on property taxation in the US shows that erroneous valuation makes the property tax regressive, in the sense that expensive houses are assessed at a lower percentage of their market value than less expensive houses (Berry, 2021; Avenancio-León and Howard, 2022; Amornsiripanitch, 2022).[2]

More specifically, we draw attention to the need for a well-functioning housing valuation scheme for taxation of wealth. This is its main usage in the Norwegian case. The annual wealth tax is back on the policy agenda in several other countries as well, see for instance Advani, Chamberlain, and Summers (2020) for the UK and Saez and Zucman (2019) for the US. As noted by Piketty and Zucman (2014), there has been a rise in the capital/income ratio, with housing contributing to this trend.[3] This means that the question of credible asset valuations are brought up. Compliance costs on the taxpayer and increased administrative costs were emphasized by Boadway, Chamberlain, and Emmerson (2010) in their report on taxation of wealth for the Mirrlees Review, and accordingly, they point to valuation information problems as a main obstacle when they conclude against recommending an annual wealth tax for the UK. Norway is one of the few OECD countries that still levy a comprehensive annual tax on net wealth (Scheuer and Slemrod, 2021; Thoresen, Ring, Nygård, and Epland, 2022), and we believe that a reasonably well-functioning housing valuation system has been essential for its survival.

The remainder of the paper is organized as follows. In Section 2 we review different methods to establish house values for tax objectives. In Section 3 we present the background for the initiatives in Norway to establish a valuation system based on hedonic regressions. Furthermore, Section 4 describes the model in current use and presents a model which could replace it in the near future. The empirical fit of the existing and new model is presented in Section 5. Section 6 concludes the paper.

[2]In practice, a number of exemptions and tax reliefs might still make property taxes in the US progressive (Ihlanfeldt and Rodgers, 2022).

[3]See also Cowell, Karagiannaki, and McKnight (2018) and Balestra and Tonkin (2018) on the distribution of housing wealth. Both studies find that housing wealth is very unequally distributed among households in developed countries.

2 Housing valuation methodology

2.1 What, how, and who

According to Almy (2014), there are three important questions to be asked in a discussion of valuation of immovable property for tax purposes: What is valued? How is it valued? And who values?

With respect to the first question, the main choice is whether to tax land, building structure, or both. The Norwegian annual wealth tax, which is the main driver of the need for valuation in the Norwegian context, includes a joint valuation of land and buildings. While there is theoretically a good case for taxing land separately, at higher rates (Bonnet, Chapelle, Trannoy, and Wasmer, 2021), it is much easier to obtain values for the combination of land and building structures, than to value them separately; see Mirrlees et al. (2011), ch. 16, for an overview.

Under the heading "How is it valued?", an important question is which measure of value one uses for the tax base. Dependent on which tax bases the housing values are used for, there are several options for valuations (Almy, 2014). In the Norwegian case, in which housing values enter into taxable wealth, primary housing is valued at 25 percent of the market value of the dwelling. Secondary housing enters the tax base at a higher value, increasing from 40 percent to 100 percent of market value over the period 2010 – 2023. This means that an annual market value is first imputed for all properties, from which the tax value is calculated.

The main topic of the present study – valuation and data collection methodology – is categorized under "Who values?" in Almy (2014). There are two main methods: data collection based on the work of inspectors from the tax administration and a system based on self-reporting by the taxpayers. In addition, information can be required in connection with an event, such as transfer of ownership. With digital developments, there is increased attention on mass valuation systems, also with the ambition to predict the value of single properties. In the following we briefly describe how different countries approach this task, before focusing on how the Norwegian version of such a system has been established and works.

2.2 Valuation schemes elsewhere

Many countries impose various taxes on immovable property, such as housing, making the requirement for establishing housing values widespread. For a systematic classification of valuation methods, see Almy (2014). Here, we go more in depth on the valuations used by a few selected countries.

Considering the most common situation, where the tax basis is supposed to reflect the market value of the property, there are two separate problems in the valuation. First, how to make an accurate valuation of all taxable properties, and second, how to keep the valuation up to date as housing values fluctuate over time.

Sweden uses a system of tables that values a "standard" house, and then adjusts value based on a combination of housing size, standard, location, and other factors (Skatteverket, 2023). These adjustment factors are likely set in relation to observed market values, though the exact procedure is not discussed.[4] The valuations are updated every three years.

[4] Swedish property taxes are set as a proportion of tax valuation, but capped at a low value, meaning that for many properties, the exact tax valuation does not influence the payable tax (Skatteverket, 2024).

Denmark is in the process of developing an automatic valuation system based on nearest neighbor transactions with adjustments for hedonic characteristics (Vurderingsportalen, 2024). The project started in 2014, is years late and still unfinished, and establishing it is estimated to have cost 4 billion DKK[5] (Rigsrevisionen, 2024). Before this, Denmark used to value housing based on local average prices, with manual adjustments (Rigsrevisionen, 2013). In addition to a tax on property value, there is a separate land tax in Denmark.

In Finland, land and buildings are subject to separate real estate taxes, with values set by the government according to a number of characteristics. For land, geographic location is important, but valuations are quite old. For buildings, geographic location is irrelevant, and the key concept is the replacement value, which is based on estimated construction costs less depreciation (Lyytikäinen, 2012). A new system for land and building valuation is being developed, to better reflect market values (Valtiovarainministeriö, 2024).

The local property valuation system of the Netherlands is based on so-called WOZ values, set at the local level, and meant to reflect market values. Obtaining these values may involve the local authorities using regression models with similarities to the Norwegian framework presented here (Netherlands Council for Real Estate Assessment, 2014; Francke, Hans, Korevaar, and van Bekkum, 2023).

In the United States, local governments such as cities and counties are responsible for levying property taxes, which vary across the country. Nationally, property taxes contribute to over 30 percent of total local tax revenue. The assessment of property taxes can be based on the property's value at the time of acquisition or can fluctuate based on the assessed market value determined by government officials (Harris and More, 2013).[6] There is evidence that property tax valuations do not closely follow market value fluctuations, instead reacting to local budget conditions (Chen and Cohen, 2024). Thus, there appears to be limited adoption of regression-based automated valuation models. This observation is reinforced by a study conducted by Bidanset and Rakow (2022) examining the utilization of automated valuation models in government assessment offices. Only 16 percent of the surveyed employees at assessment offices indicated the use of some form of regression-based model in their assessments.[7]

The use of outdated housing valuations is common in many countries. The English (and Scottish) council tax is based on housing valuations from 1991 (Mirrlees et al., 2011), whereas German tax authorities use property values from 1964 to establish a tax base for their property tax (Löffler and Siegloch, 2021). Such schemes will result in unequal tax burdens if price growth has been different between different regions or housing segments. Absent a system that automatically updates property values, there is a significant political cost of updating property valuations. As housing prices mostly have been increasing over the last decades, more accurate values would likely result in higher taxes.

While some other countries use methods that, like the Norwegian system, try to reflect updated market values of property, the simplicity of the Norwegian system is unusual. The Norwegian system is also transparent, in that the functional form of the hedonic regression model, as well as

[5] Around USD 600 million or EUR 530 million (2020-values).

[6] See also the web page of The Lincoln Institute of Land Policy, which provides a state-by-state overview of the property taxes in the US.

[7] The study also included employees at government offices outside the US. However, 93 percent of respondents are from jurisdictions in the US.

the parameter values of the model, are published.[8] The structure and transparency of the system also reduces the scope for opportunistic behavior in valuation, while discretionary systems may allow tax assessors to influence valuation for private gain, as shown in Chen and Cohen (2024).

3 A history of different valuation systems

3.1 A problematic valuation system

The demand for housing valuations is a function of the tax system. At the time being, residential property values enter into the tax bases of two types of taxes in Norway – the annual tax on net wealth and the annual tax on property. The two other forms of taxation that included housing valuations, the tax on imputed housing rent and the inheritance tax, were abolished in 2005 and 2014, respectively. The wealth tax is administered by the the central government, whereas the property tax is a municipal tax, leaving the decision to levy it and which type of properties to include in the tax base to the local government. At the time of writing (2024), approximately 90 percent of the municipalities have a type of property tax. However, although many municipalities use the central government valuation system, many of them instead do their own assessments, which are based on personal valuations by an inspector.

A key reason for introducing a completely new housing valuation system in 2010 was the recognition that the previous system was highly problematic, with significant weaknesses in terms of tax legitimacy. It delivered house values which were far from market values, and prediction errors differed across geography and housing characteristics (Ministry of Finance, 1996). The long-standing guidelines from the tax authorities suggested that the tax value for new houses should be set to 20–30 percent of the total cost of the dwelling, which included both land and construction costs. However, as these values were kept nominally unaltered over many years, several dwellings had values which were substantially below the 20–30 percent range. For example, it was well-known that apartments of the up-market area Frogner in Oslo, established in the 19th century, had tax values far below this level.

Although there were several initiatives to restructure the system, it was difficult to reform. One important reason was that a change in the valuation would have significant implications for taxation and the distribution of tax burdens, which meant that it was difficult to judge a new system in isolation from the tax consequences.

The most powerful initiative to establish a proper and a more legitimate system came through a white paper in 1996 (Ministry of Finance, 1996). At that time, housing values entered into the tax base of four different taxes: in addition to the wealth tax and the local property tax, which are still in place, housing values were part of the tax base for the inheritance tax and the personal income tax included taxation of imputed rent. The government suggested a new appraisal system based on the following main components. First, the value of the property was determined by the production value of the building plus market value for the land. Second, the production value was based on a square-meter price dependent on the type of building (house, large and small apartment buildings) and geographical location. There were reductions in value for depreciation, age and being outdated. Third, for land value, the country was divided into approximately 10 price zones, with two different

[8] As an example, Takle and Medby (2022) shows the details of the 2021-model.

rates, dependent on how many housing units belonged to the piece of land. Fourth, individual corrections could be made if the difference between the calculated value and the market value was more than 30 percent. Fifth, recreational housing (such as cabins) was supposed to be valued by the same system.

Importantly, with respect to the tax implementation, the intention was to let housing be more leniently taxed than other capital items, and to keep it low. To shelter owners of standard dwellings, the new valuation system was paired with generous deduction schemes for the taxes in question (Sørvoll, 2011).

This system was never implemented, primarily because it met substantial political opposition. The rather complicated system was susceptible to political and technical disagreements along several dimensions. Moreover, in addition to being costly to implement, there were concerns about the system's ability to ensure fair taxation (Sørvoll, 2011). In this perspective, a less complicated scheme could be less vulnerable to political controversies. Next, we discuss a much simpler, and actually realized procedure to assess housing values, through a scheme based on hedonic regressions.

3.2 Hedonic valuation

Transitioning into the 2000s, there remained a persistent call for a fair housing valuation system. Building upon the lessons learned from the protracted deliberations of the 1990s, the center-left government (Red-Green Coalition) in 2009 pursued a more streamlined political approach: simply incorporating a new valuation system into the 2010 budget (Ministry of Finance, 2009). The change was explicitly connected to making tax valuations better reflect market values. Two critical factors proved essential for the successful adoption of the new scheme in 2010: high quality data and a wealth tax reform to make it generally acceptable.

With respect to data, the establishment of a new valuation system came as a result of a long-standing work to improve data on housing. To create the hedonic model, it was important to collect detailed information on transactions. This was done by retrieving information from FINN.no, which is by far the largest marketplace for housing transactions in Norway. This digital market place was established back in 2000 and today covers about three-quarters of the total housing transactions. However, the degree of coverage depends on the price level and geographical area. The most expensive houses, worth 30 million NOK (3.2 million US dollars or 3 million euros)[9] or more have a coverage of less than 30 percent, and the least expensive houses (below 1 million NOK) have a coverage of less than 20 percent. Moreover, urban areas have a significantly higher degree of coverage, i.e., for Oslo the data from FINN.no cover more than 96 percent of the total transactions (Takle and Medby, 2020). By collecting data for several years, one obtains a sufficiently large number of observations of transactions, and market values are obtained by using predicted values from hedonic regression analysis. More details on the hedonic model are provided in Section 4.[10]

For predicting tax values of housing units that had not been transacted over the last years, it was necessary to obtain information on the size and age of each housing unit. It was also essential to associate each housing unit in the country with an owner. A property register already existed within

[9] According to 2020 exchange rates: USD 1 equals NOK 9.4, EUR 1 equals NOK 10.1.

[10] The model has been documented in several reports, such that Thomassen and Melby (2009), Takle and Medby (2022), and Bø, Medby, Nygård, and Takle (2022); all of them are in Norwegian.

the tax administration, as housing had already been subject to taxation. If the taxpayers find the predicted tax value to be too high compared to the actual market value, a correction claim can be submitted to the Norwegian Tax Administration.[11]

The second key component in order to implement a regression-based system was to make it acceptable amongst the citizens. Thus, its introduction was combined with a reform of the wealth tax scheme. The main purpose of a new valuation system was taxation of annual net wealth; see Thoresen et al. (2022) for a description of the Norwegian wealth tax and its historical development. As the more precise housing valuations would increase the wealth tax for a large share of housing owners, there were concerns that the change could be unpopular. To make sure that most taxpayers would instead experience a tax relief in the same year as the new valuation system was introduced (2010), the wealth tax was reduced for most taxpayers. While the new housing valuation model in isolation was estimated to *increase* wealth taxes by 1.3 billion NOK, higher minimum deductions meant that the net effect on wealth taxation was a *decrease* of 760 million NOK (Ministry of Finance, 2009).

Over time, however, the model predictions have caused some controversy – in particular very expensive houses have been shown to be under-valued by the method. Moreover, the model, which uses municipalities as the main geographical unit, is not well suited to handle changes that have occurred in the municipality structure. Further exploration of the existing model is presented in Section 4. As a result, a new revised model has been developed (Bø and Nygård, 2023), which we also present in the following.

The tax valuations described here can also be used for the municipal property tax. Local authorities decide both whether to impose a municipal property tax or not, and which valuation system to use. Given that the local property tax is non-mandatory and varies across municipalities, we focus on the use of the valuation schemes for the annual wealth tax in the following. It is worth noting though, that one large municipality has stated weaknesses in the current valuation model as a reason for switching to a self-developed valuation method (Trondheim kommune, 2023).

4 Hedonic regression model

4.1 The existing model

Based on hedonic regressions on data from housing transactions, the effects of certain observable characteristics on housing values is estimated. These estimates are then used to predict the housing values of all existing residential properties in Norway. An important prerequisite for the valuation model was access to adequate data sources. The hedonic model uses data on transaction prices, housing size, age and geographical location (i.e., in which municipality the property is situated).

The most important data component is data on values. Transaction data were gathered from FINN.no, which was (and is) the largest marketplace for housing transactions in Norway. In addition to transaction price and time, the data from FINN.no also include information on housing size and housing age. There is a rich set of additional information in the data, such as number of bathrooms, bedrooms, garages, the existence of fireplaces and balconies. However, the potential use of this

[11] Specifically, if the taxpayer can demonstrate that the model values the property at more than 120 percent of its market value, the model's valuation can be challenged.

additional information is limited by the need for the same information for non-transacted properties, for which housing values are to be predicted.

Housing age and size were available for all properties through administrative sources, but the accuracy of the information was uncertain. Thus, at the introduction of the system (at tax filing in 2011), all housing owners were required to supply housing size and housing age for their properties. The other variables were already available in administrative registers.

The market based tax value of housing is estimated using a simple hedonic regression framework. The following regression is run separately for 3 different housing types and a number of different geographic regions:

$$\ln price = \alpha + \beta \ln size + \tau dense + \lambda year + \theta age + \delta zone + \varepsilon, \tag{1}$$

where the log price per square meter is the dependent variable. The independent variables are the *size* (in square meters); *dense*, which is a dummy variable for housing units in areas classified as densely built; the transaction *year*, fixed effects for year sold; the *age* of the housing unit (classified into four categories); and geographical *zones* within each region. The housing types are single-family houses, semi-detached houses and apartments. There are 21 different regions[12] for single-family houses, and 11 regions for the two other housing types (reflecting the prevalence of single family houses in rural regions). Thus, the regression in Equation (1) is run separately for 43 combinations of region and housing type. The regressions use housing transaction data from the last 10 years. In 2010, only six years were used, but the period was expanded as more data became available.

Using the estimated parameter values from Equation (1), housing values are then predicted for the whole Norwegian housing stock, by the equation:

$$V = \exp(\hat{\alpha} + \hat{\tau} dense + \hat{\lambda} year_T + \hat{\theta} age + \hat{\delta} zone + \frac{\hat{\sigma}^2}{2}) size^{(1+\hat{\beta})}, \tag{2}$$

where V is the predicted housing value, and $\hat{\sigma}$ is the estimated standard deviation of the error term of Equation (1). The values are predicted for the most recent year, denoted $year_T$. Using the tax year 2020 as an example, this means that in early 2021, transaction data for 2011–2020 were used to estimate predicted housing values for all houses in 2020. This value is included in the pre-filled tax filings that tax authorities sent to households around March 2021.

The model performs best in areas with dense population, where there are many observed transactions (Takle and Medby, 2022). Likely related, the models has a worse performance for single-family houses, which are spread across the whole country, than for semi-detached houses and apartments.

There are however a number of challenges with the current model. Any hedonic model will have difficulties in valuing unusual properties, e.g. houses with uncommonly high or low standard. The problem of unobservable factors is likely exacerbated by the current model only including a very limited number of explanatory variables. In the public debate in Norway, there has been a particular focus on the current model underestimating the value of expensive houses. Thus, a demand for an improved model has emerged, which we will return to next.

[12] Originally 22, reduced with the merger of Nord- and Sør-Trøndelag in 2018.

4.2 A revised model – clustering with machine learning

The need for improvement of the current housing valuation model has been noted at the governmental level (Norwegian Government, 2021). In general, the demand for improvements results both from observed weaknesses of the current model (as described in the last section), and because municipality mergers worsened the predictive power of the model. A way of predicting Norwegian housing valuations by utilizing the geographical information more flexibly is discussed in Bø and Nygård (2023).[13] We will present the improved model in this section.[14]

Other ways to improve a hedonic regression by incorporating more information on geographic structure are found in Hill and Scholz (2018) and Ahlfeldt, Heblich, and Seidel (2023). Further, there is a growing literature on valuation models utilizing machine learning (Kok, Koponen, and Martínez-Barbosa, 2017; Fagereng, Holm, and Torstensen, 2020; Stang, Krämer, Nagl, and Schäfers, 2023), although to the best of our knowledge, no studies have yet explored the use of machine learning models for tax valuation purposes. One potential challenge of using machine learning models for tax valuation is the lack of transparency in how these models operate. This can be problematic if there is a need to explain a specific tax value to a taxpayer. However, recent advancements have been made in improving the explainability of machine learning models (Krämer, Nagl, Stang, and Schäfers, 2023).

The revised model is designed to be close to the current model in terms of specification and data requirements, to make a transition as simple as possible. The main difference is going from using municipalities to basic statistical units (BSUs) as the geographical unit. BSUs are the smallest geographical units used by Statistics Norway for statistical purposes. Norway is split into around 14,000 BSUs. Each BSU is designed to be relatively homogeneous in terms of housing structure, often delineated by natural borders such as rivers or major roads.

In the existing model, the calculation areas and price zones were manually constructed compositions of counties and municipalities, based on geographic closeness and housing price levels. Using this method to group 14,000 BSUs is not practicable. Thus, the new model uses a simple machine learning method to group the BSUs.

Specifically, we use a machine learning algorithm called a Decision tree to group similar BSUs.[15] The algorithm groups BSUs into calculation areas that minimize the squared error between observed square meter prices and average square meter prices within the calculation areas. Similarity of BSUs is thus measured by price per square meter (dependent on housing type). The dependent variables are geographical location (measured as the average latitude and longitude of transactions within the BSU), municipality and county. The algorithm recursively partitions the dataset into groups, based on the dependent variables. The partitioning continues until further partition would create calculation areas with fewer observations than the chosen minimum (which is 500).[16] The grouping of BSUs into calculation areas is done separately for three different housing types (single-family houses, semi-detached houses and apartments), as in the existing model.

The new procedure results in around 1,200 different calculation groups (combination of housing type and BSU based geographical area, containing at least 500 observations). In the larger cities,

[13]Which is an extension of Bø et al. (2022).

[14]The presentation closely follows Bø and Nygård (2023).

[15]See Mullainathan and Spiess (2017) for an introduction to machine learning in economics.

[16]For an introduction to partitioning methods, including decision trees, see Strobl, Malley, and Tutz (2009).

there may be enough inhabitants and transactions that the calculation groups consist of only one or a few BSUs, while in rural areas, calculation groups might be rather large geographical entities.

Then, a hedonic regression is run for each calculation group:

$$\ln price = \alpha + \beta \ln size + \lambda year + \theta age + \delta BSU + \varepsilon, \quad (3)$$

where (as above) the log price per square meter is the dependent variable. The independent variables are the *size* (in square meters); the transaction *year*; the *age* of the housing unit (classified into four categories); and fixed effects for each *BSU*. The specification of Equation (3) is by design close to the specification of Equation (1), though the *dense* dummy, which is defined at a municipal level, is unnecessary. The *BSU* fixed effect replaces both the *dense* dummy and the geographical *zone* fixed effect. A transformation similar to Equation (2) transfers the predicted log value into the predicted housing price. This method requires no additional information beyond what is already used in the existing model, except for the coordinates of the housing units.

By having more and often much smaller calculation groups, this model is able to better utilize local price information in urban areas, where transactions data are abundant, than the existing model.

5 Empirical fit of the models

In this section, we look at the empirical fit of the tax valuation models, both the existing and the revised models. The empirical fit is measured as the difference between predicted price and the transaction prices for the subsample of properties that have been transacted over the 10 year period 2011–2020.[17] Mainly, we will be presenting figures of the distribution of the absolute percentage deviation of the prediction errors.

In Table 1, we present results for the whole sample. We compare four different measures: the average and median of the absolute prediction error, the standard deviation of the prediction error, and the R^2 of the prediction model. All four measures show markedly improved fit for the revised model. The first three measures decrease by 25–30 percent, while the R^2 clearly increases. It is also worth noting that the existing model, even though very simple, has a good fit, with an R^2 of 0.83.

Table 1: Model performance of the existing and revised models

Model	Existing	Revised
Average absolute error (percent)	15.58	11.29
Median absolute error (percent)	11.48	8.11
Standard deviation	22.44	16.72
R^2	0.833	0.905

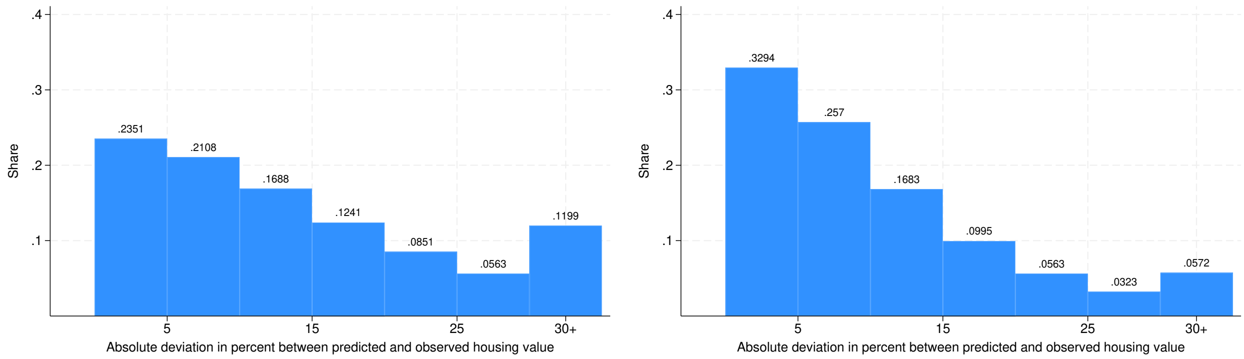
Notes: The table describes model performance by four different measures, comparing predicted prices to realized prices from transaction data.

Figure 1 shows the distribution of the absolute prediction errors. First, in Figure 1A, the absolute

[17] Transactions where buyers and sellers are closely related, as well as forced sales, are excluded.

Figure 1: Distribution of prediction errors

(A) Existing model | **(B)** Revised model



Notes: The distribution of predictions errors from two models: The existing model (panel (A)) and the revised model (panel (B)). Prediction errors are shown as the absolute percentage deviation between predicted and observed housing values for the full set of transacted houses.

percentage deviation between predicted values from the existing model and observed transaction values is depicted. Figure 1B show the same differences between the transaction values and the revised model. The revised model shows clear improvements in the empirical fit. Approximately 60 percent of the predictions from the existing model fall within a 15 percent deviation of the observed values, whereas this increases to around 75 percent for the revised model. In the existing model, 24 percent of observations have a prediction error of less than 5 percent, 45 percent less than 10 percent, while 18 percent have errors of 25 percent or more. This compares to 33 percent, 59 percent and 9 percent for the revised model.[18]

It is acknowledged that the current model does not produce good predictions for the most expensive houses. To see if the revised model improves these predictions, we next look at the most expensive houses in Oslo. Table 2 shows measures of fit for housing transactions in Oslo with a transaction price of 10 million NOK or more.[19] The existing model does clearly worse for this subsample, with an R^2 of only 0.42.[20] The revised model improves the fit by 40 percent.

Table 2: Model performance of the existing and revised models. Expensive houses

Model	Existing	Revised
Average absolute error (percent)	17.49	12.87
Median absolute error (percent)	15.44	10.50
Standard deviation	18.22	15.38
R^2	0.415	0.590

Notes: The table describes model performance for expensive housing in the Oslo-area (transaction price of 10 million NOK or more) by four different measures, comparing predicted prices to realized prices from transaction data.

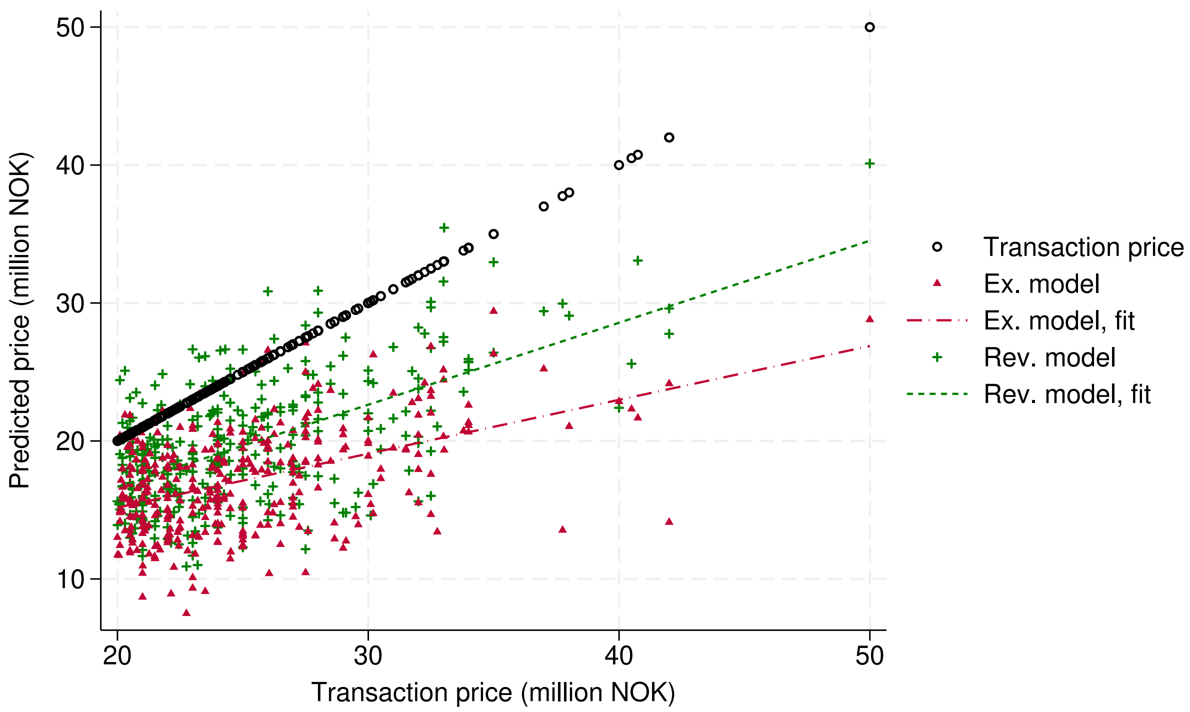
[18]Further analyses in Bø and Nygård (2023) shows that the revised model improves the fit for all regions, building types and price groups.

[19]Around USD 1.1 million or EUR 1 million (2020).

[20]It is obvious that any model will have a worse fit for the small subsample of the most expensive transacted houses. Still, the revised model shows that a better fit is possible.

To assess the performance of the prediction models for the most expensive houses, Figure 2 shows the transaction price and the predicted prices from the two different models for the 369 houses that were sold for more than 20 million NOK over the period (2011–2020).[21] The predicted prices are almost all below the black circles, i.e., below the transaction prices, particularly for the most expensive houses. However, the green crosses of the revised model are generally closer to the transaction price than the red triangles of the existing model (as shown by the linear regression fit lines).

Figure 2: Predicted prices for expensive houses



Notes: Comparison of predictions from to the existing model (red triangles) and the revised model (green crosses) to transaction prices (black circles) for expensive houses (sold for more than 20 million NOK over the period 2011–2020). Fitted lines obtained by linear regression.

5.1 Revenue and distribution

The new valuation system has implications for both wealth tax revenue and distributional effects of the wealth tax. We use the tax-benefit model LOTTE (Jia, Larsen, Lian, Nesbakken, Nygård, Thoresen, and Vattø, 2024) to obtain information on effects of going from the existing system to the new system. We find that the tax revenue from the wealth tax increases by 1.6 percent, which is caused by higher valuation of expensive houses. While the average predicted housing value is equal in both models, expensive houses get higher values, and they are to a large extent owned by people who pay wealth taxes. Houses that get lower valuations, often cheap housing in rural areas, affect tax revenue less, as their owners often do not pay any wealth tax. There are 300,000 persons who get an increase in wealth tax, while 225,000 persons get a decrease.

[21] The set of housing transactions over 10 million NOK, as presented in Table 2, becomes too large to effectively visualize in this way.

To assess the distributional effects of the revised model, we first focus on individuals impacted by the changes – those whose wealth tax payments are affected by the new model. In Table 3, these individuals are ranked by gross income, and we estimate the change in tax payments for each income decile using the tax-benefit model. Overall, the increases in tax payments are relatively modest, reaching 1,224 NOK for decile 10 and averaging 532 NOK across all deciles. When we express the change in tax payments as a percentage of gross income, the new model appears to make the wealth tax slightly more regressive for these individuals. The most notable effect is observed in decile 1, where the change in tax payments is the highest relative to income. This is partly due to the presence of wealthy individuals in decile 1 who have low or negative incomes as a result of significant capital losses.

When examining the effects on the entire population, the overall picture changes, see Table 4. Now, the introduction of the revised model no longer makes the wealth tax more regressive. The relative impact on individuals in higher income deciles increases, while it decreases in low income deciles. This is because few people with low incomes pay any wealth tax. Additionally, as anticipated, the average change in tax payments for the whole population is very small, amounting to just 65 NOK, or 0.013 percent of gross income. The Gini index of disposable (post-tax) income is virtually unchanged.[22]

Table 3: Existing versus revised model for 2020. Individuals experiencing change in tax payments ranked by gross income

Decile	Gross income	Change in tax payments	As share of income (%)
1	188,639	371	0.197
2	304,634	264	0.087
3	369,589	303	0.082
4	430,995	377	0.087
5	501,306	408	0.081
6	588,017	473	0.080
7	695,056	517	0.074
8	851,473	634	0.074
9	1,134,675	745	0.066
10	3,024,474	1224	0.040
All	808,886	532	0.066

Notes: Estimates are averages in NOK obtained by the use of the tax-benefit model LOTTE-Skatt. Change as percentage share of income is calculated as the average of change divided by the average of income. The table includes individuals 17 years and older experiencing change in tax payments.

6 Conclusion

Lack of information on the values of taxable items represents an obstacle in the design of well-functioning tax schemes. Both for tax schemes that involve return to housing as part of the income

[22]If imputed housing income had been included in the income definition, the revision would likely have made the revision progressive. That result in the context of a property tax increase, as well as a further discussion on the role of imputed housing income, can be found in (Bø, 2020).

Table 4: Existing versus revised model for 2020. All individuals ranked by gross income

Decile	Gross income	Change in tax payments	As share of income(%)
1	41,994	14	0.034
2	187,269	21	0.011
3	270,464	23	0.009
4	333,858	33	0.010
5	399,554	46	0.012
6	467,034	46	0.010
7	539,054	49	0.009
8	623,770	57	0.009
9	759,918	88	0.012
10	1,442,334	269	0.019
All	506,525	65	0.013

Notes: Estimates are averages in NOK obtained by the use of the tax-benefit model LOTTE-Skatt. Change as percentage share of income is calculated as the average of change divided by the average of income. The table includes all individuals 17 years and older.

tax and housing wealth as base for the property tax or other types of wealth taxation, it is essential for tax administrations to obtain accurate housing values. In real world settings, this turns out to represent a common challenge.

The main message of this paper is that estimating housing values does not need to be administrative burdensome nor involve high costs. We argue that employing a hedonic valuation scheme is a practical alternative with low costs. This paper demonstrates how such a scheme can be implemented in practice. To our knowledge, this valuation arrangement, i.e., establishing nationwide tax bases from econometric evidence, is rather unique.

The procedure is illustrated with the Norwegian case, where a regression-based scheme for housing valuation has been in place since 2010. A market value for each dwelling in Norway is predicted based on estimates from hedonic regressions, with the values updated yearly. The scheme has a limited need for data, and is easy to update. We also present a suggested redevelopment for the valuation system, where information at a more granular level is used. We show that model performance is clearly improved by using the "basic statistical unit" (grunnkrets) as the unit of analysis instead of employing data at the municipality level. However, according to our judgment, both methods work sufficiently well to be used in practice. While much of the literature on automatic valuation of properties focus on one class of housing, or a single city, the models described here works well for all different residential housing types in the whole country of Norway.

There is a question about how applicable this method would be outside of Norway. Norway is a small country, with 5.5 million inhabitants, but geographically, it is relatively large. There is a lot of variation both in the geographic features and in the density of Norwegian settlements. In particular, it is interesting to see that the rather simple model also has a reasonably good fit in rural areas where there are few comparable transactions. In more densely populated countries than Norway, a similar model would likely have an even better fit. The BSUs used in the clustering of the revised model are well-suited for the purpose as they are designed to be homogeneous, but the procedure could likely work with any small, geographical unit such as a zip-code. The availability of sufficient data is a requirement, but as illustrated here, a good model can be made without extensive use of housing unit microdata.

It follows that evidence obtained from econometric analysis can work well to establish a tax base – in our case as a base for taxation of annual net wealth. Such a valuation scheme is worth considering for tax administrations of other countries, either to derive housing values or to improve deficient valuations on other items that enter the tax base. Using econometric methods for administrative tax purposes represents a promising procedure for future information enhancement. The ongoing development of machine learning methods should facilitate further accuracy in automatic property valuation, either integrated in a hedonic framework, or as full machine learning models.

There are also valuable insights to be learned regarding the implementation of the Norwegian valuation system. The new housing values introduced in 2010, more closely reflecting real market values, would in isolation have led to an significant increase in the wealth tax. It is our belief that a significant factor contributing to the widespread acceptance of the new framework in 2010 was the concurrent increase in minimum deduction, which led to a net reduction in the wealth tax.

References

Advani, A., E. Chamberlain, and A. Summers (2020). A wealth tax for the UK. Wealth Tax Commission. Final Report, United Kingdom.

Ahlfeldt, G. M., S. Heblich, and T. Seidel (2023). Micro-geographic property price and rent indices. *Regional Science and Urban Economics 98*, 103836.

Almy, R. (2014). Valuation and assessment of immovable property. OECD Working Papers on Fiscal Federalism No. 19, OECD Publishing, Paris.

Amornsiripanitch, N. (2022). Why are residential property tax rates regressive? Philadelphia Fed Working Paper 22-02, Federal Reserve Bank of Philadelphia Supervision, Regulation, and Credit Department.

Avenancio-León, C. F. and T. Howard (2022). The assessment gap: Racial inequalities in property taxation. *The Quarterly Journal of Economics 137*(3), 1383–1434.

Balestra, C. and R. Tonkin (2018). Inequalities in household wealth across OECD countries: Evidence from the OECD Wealth Distribution Database. OECD Statistics Working Papers No. 2018/01, OECD Publishing, Paris.

Berry, C. R. (2021). Reassessing the property tax. Unpublished manuscript.

Bidanset, P. and R. Rakow (2022). 2019 survey on the use of automated valuation models (avms) in government assessment offices: An analysis of avm use, acceptance, and barriers to more widespread implementation. Working paper WP22PB1, Lincoln Institute of Land Policy.

Boadway, R., E. Chamberlain, and C. Emmerson (2010). Taxation of wealth and wealth transfers. In S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, G. M., P. Johnson, G. Myles, and J. Poterba (Eds.), *Dimensions of Tax Design: The Mirrlees Review*, pp. 737–814. Oxford University Press.

Bonnet, O., G. Chapelle, A. Trannoy, and E. Wasmer (2021). Land is back, it should be taxed, it can be taxed. *European Economic Review 134*, 103696.

Bø, E. E. (2020). Taxation of housing: Killing several birds with one stone. *Review of Income and Wealth 66*(3), 534–557.

Bø, E. E., P. Medby, O. E. Nygård, and M. Takle (2022). Hedonisk verdsettelse av bolig med fleksibel geografisk inndeling. *Tidsskrift for boligforskning 5*(2), 70–87 (in Norwegian).

Bø, E. E. and O. E. Nygård (2023). Nye beregningsområder i boligformuemodellen. Notater 2023/3, Statistics Norway (in Norwegian).

Chen, H. and L. Cohen (2024). Assessing assessors. NBER Working Paper 33238, NBER, Camebridge, MA.

Cowell, F., E. Karagiannaki, and A. McKnight (2018). Accounting for cross-country differences in wealth inequality. *Review of Income and Wealth 64*(2), 332–356.

Eckert, J. K. (2008). Computer-assisted mass appraisal options for transitional and developing countries. In R. Bahl, J. Martinez-Vasquez, and J. Youngman (Eds.), *Making the Property Tax Work*, pp. 737–814. Cambridge, Massachusetts: Lincoln Institute of Land Policy.

Fagereng, A., M. B. Holm, and K. N. Torstensen (2020). Housing wealth in Norway, 1993–2015. *Journal of Economic and Social Measurement 45*(1), 65–81.

Figari, F., A. Paulus, H. Sutherland, P. Tsakloglou, G. Verbist, and F. Zantomio (2017). Removing homeownership bias in taxation: The distributional effects of including net imputed rent in taxable income. *Fiscal Studies 38*(4), 525–557.

Francke, M., L. Hans, M. Korevaar, and S. van Bekkum (2023). Buy-to-live vs. buy-to-let: The impact of real estate investors on housing costs and neighborhoods. Unpublished manuscript.

Grover, R. (2016). Mass valuations. *Journal of Property Investment & Finance 34*(2), 191–204.

Harris, B. H. and B. D. More (2013). Residential property taxes in the United States. https://www.taxpolicycenter.org/publications/residential-property-taxes-united-states/full, Tax Policy Center.

Hill, R. J. and M. Scholz (2018). Can geospatial data improve house price indexes? A hedonic imputation approach with splines. *Review of Income and Wealth 64*(4), 737–756.

Ihlanfeldt, K. and L. P. Rodgers (2022). Homestead exemptions, heterogeneous assessment, and property tax progressivity. *National Tax Journal 75*(1), 7–31.

Jia, Z., B. Larsen, B. Lian, R. Nesbakken, O. E. Nygård, T. O. Thoresen, and T. E. Vattø (2024). The LOTTE system of tax microsimulation models. *International Journal of Microsimulation*, forthcoming.

Kok, N., E.-L. Koponen, and C. A. Martínez-Barbosa (2017). Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management 43*(6), 202–211.

Krämer, B., C. Nagl, M. Stang, and W. Schäfers (2023). Explainable AI in a real estate context–Exploring the determinants of residential real estate values. *Journal of Housing Research 32*(2), 204–245.

Löffler, M. and S. Siegloch (2021). Welfare effects of property taxation. Discussion Paper 21-026, ZEW - Centre for European Economic Research.

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.

Ministry of Finance (1996). Boligtaksering og prinsipper for boligbeskatning. St meld nr 45 (1995-96), Government of Norway (in Norwegian).

Ministry of Finance (2009). Prop. 1 S for budsjettåret 2010: Skatte-, avgifts- og tollvedtak. Government of Norway (in Norwegian).

Mirrlees, J. et al. (2011). *Tax by design: The Mirrlees review*. Oxford University Press.

Mullainathan, S. and J. Spiess (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

Netherlands Council for Real Estate Assessment (2014). Property valuation and taxation in The Netherlands. Report.

Norregaard, J. (2013). Taxing immovable property: Revenue potential and implementation challenges. International Monetary Fund Working Paper 13/129, IMF, Washington.

Norwegian Government (2021). Hurdalsplattformen. https://www.regjeringen.no/no/dokumenter/hurdalsplattformen/id2877252/, Government of Norway (in Norwegian).

Piketty, T. and G. Zucman (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics 129*(3), 1255–1310.

Rigsrevisionen (2013). Beretning til statsrevisorerne om den offentlige ejendomsvurdering. Rigsrevisionen (in Danish).

Rigsrevisionen (2024). Skatteministeriets styring af det nye ejendomsvurderingssystem. RN 1504-24, Rigsrevisionen (in Danish).

Saez, E. and G. Zucman (2019). *The Triumph of Injustice: How the Rich Dodge Taxes and How to*

Make Them Pay. WW Norton & Company.

Scheuer, F. and J. Slemrod (2021). Taxing our wealth. *Journal of Economic Perspectives 35*(1), 207–230.

Skatteverket (2023). Skatteverkets allmänna råd om riktvärdeangivelser och grunderna för taxeringen och värdesättningen av småhusenheter vid 2024 års förenklade fastighetstaxering. SKV A 2023:8, Skattverket (in Swedish).

Skatteverket (2024). Municipal and national property tax. https://www.skatteverket.se/servicelankar/otherlanguages/inenglishengelska/individualsandemployees/declaringtaxesforindividuals/owningrealpropertyinswedenlivingabroad/municipalandnationalpropertytax.4.676f4884175c97df41923c6.html, Skatteverket.

Sørvoll, J. (2011). *Norsk boligpolitikk i forandring 1970-2010*. Oslo Metropolitan University-OsloMet: NOVA (in Norwegian).

Stang, M., B. Krämer, C. Nagl, and W. Schäfers (2023). From human business to machine learning—methods for automating real estate appraisals and their practical implications. *Zeitschrift Für Immobilienökonomie 9*(2), 81–108.

Strobl, C., J. Malley, and G. Tutz (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods 14*(4), 323.

Takle, M. and P. Medby (2020). Modell for beregning av boligformue. Notater 2020/9, Statistics Norway (in Norwegian).

Takle, M. and P. Medby (2022). Modell for beregning av boligformue. Notater 2022/8, Statistics Norway (in Norwegian).

Thomassen, A. and I. Melby (2009). Beregning av boligformue. Notater 2009/39, Statistics Norway (in Norwegian).

Thoresen, T. O., M. A. Ring, O. E. Nygård, and J. Epland (2022). A wealth tax at work. *CESifo Economic Studies 68*(4), 321–361.

Trondheim kommune (2023). Takseringsmetode bolig 2023. https://www.trondheim.kommune.no/tema/bygg-kart-og-eiendom/eiendom/eiendomsskatt/takseringsmetode-bolig-2023/, Trondheim municipality (in Norwegian).

Valtiovarainministeriö (2024). Fastighetsskattereformen. https://vm.fi/sv/fastighetsskattereformen, Valtiovarainministeriö (in Swedish).

Vurderingsportalen (2024). New public property assessment system. https://www.vurderingsportalen.dk/ejerbolig/english/, Vurderingsstyrelsen.