WORKING PAPER

DANMARKS NATIONALBANK

27 SEPTEMBER 2022 — NO. 190

# Opportunities and risks in the residential sector during a green transition: House prices, energy renovations and rising energy prices

**Alessandro T. Martinello**
*alem@nationalbanken.dk*
DANMARKS NATIONALBANK

**Niels F. Møller**
*nfm@nationalbanken.dk*
DANMARKS NATIONALBANK

**WORKING PAPER** — DANMARKS NATIONALBANK
27 SEPTEMBER 2022 — NO. 190

# Opportunities and risks in the residential sector during a green transition: House prices, energy renovations and rising energy prices

**Abstract**
Transitioning to a low-carbon economy implies both risks and opportunities in the Danish housing sector, which accounts for one fourth of Denmark's $CO_2$ emissions. We study the heterogeneous impacts on house prices of rising energy prices and energy renovations by combining micro-level data on sales and housing characteristics with data from the official mandatory energy rating reports. We find that higher energy prices reduce the prices of houses without district heating mainly in rural areas. Most renovations will not increase sales prices enough to cover the costs. Those renovations whose price effect will cover the costs have a lower-than-average impact on $CO_2$ emissions, are cheap, and typically concern houses located in and around towns and mid-sized cities and other areas with a higher population density and well-developed road networks connected to towns and cities. We conclude that while opportunities for profitable energy renovations are concentrated in these areas, transitional risks are instead associated with peripheral rural areas, where both the exposure to rising energy prices and the risk of financing renovations is highest.

**Resume**
Omstillingen til et lavemissionssamfund indebærer både risici og muligheder i den danske boligsektor, der står for en fjerdedel af Danmarks $CO_2$-udledning. Vi analyserer effekterne på prisen for en-familie huse af stigende energipriser og energirenoveringer ved at kombinere mikro-data for salgs- og boligkarakteristika med data fra energimærkningsrapporter. Vi finder, at højere energipriser reducerer salgspriserne på huse beliggende hovedsageligt i landdistrikter, og som ikke opvarmes med fjernvarme. De fleste energirenoveringer forøger ikke salgsprisen nok til at dække investeringsomkostningerne. De renoveringer, der forøger salgsprisen mere, end de koster, har typisk en lavere effekt på $CO_2$-udledningen i forhold til gennemsnittet, er ofte billigere og gælder typisk for huse beliggende i og omkring små til mellemstore provinsbyer og visse områder med højere befolkningstæthed og en veludviklet infrastruktur. Mulighederne for rentable energirenoveringer er således koncentreret i disse områder, mens risici i stedet forekommer i de mere perifere landområder, hvor både eksponeringen over for stigende energipriser og risikoen ved finansieringen af renoveringer er størst.

**Key words**
Climate; Housing finance

**JEL classification**
R30; Q40; Q50.

**Acknowledgements**
The authors would like to thank, in random order, the Danish Safety Technology Authority, Tim Thøgersen, Thomas Sangill, Thais Jensen, Marcus Ingholt, Martin Oksbjerg, Mette Petry, Ismir Mulalic, Stine Bech, Simone Bonin, Johannes Poeschl, Thomas Harr and Peter Storgaard.

CLI
MA
TE

# Climate change

Climate change is impacting society already today and will have further consequences in the future. A successful green transition will require unprecedented efforts, both in Denmark and abroad.

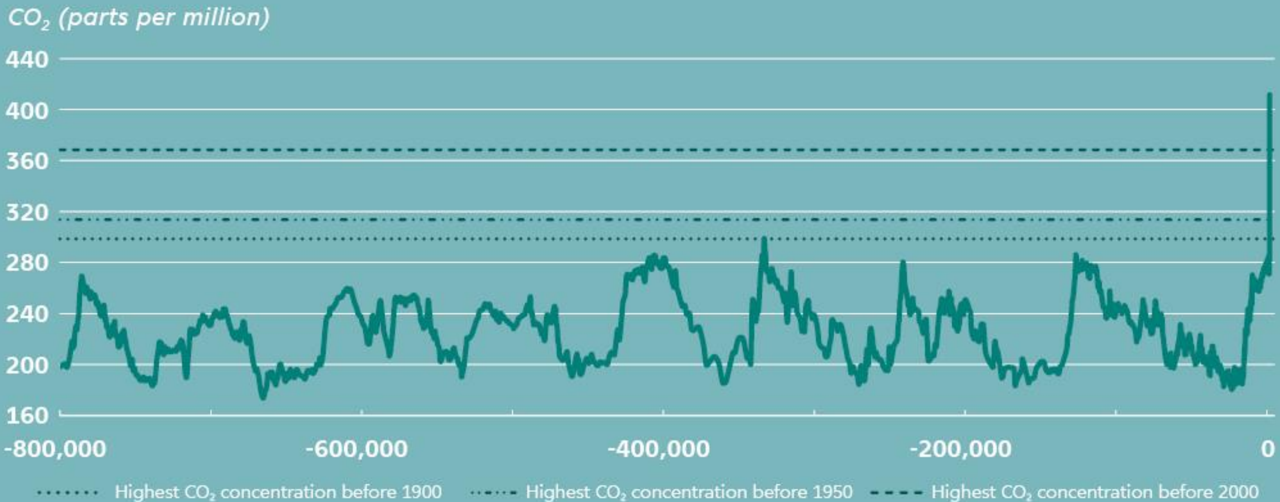
As a case in point, climate change and the transition to a greener economy will impact corporate earnings and economic activity. This may compromise price and financial stability in Denmark, which it is Danmarks Nationalbank's objective to ensure. It is therefore essential that Danmarks Nationalbank increases its knowledge of how, and by how much, the climate challenges will impact various parts of the economy.

Against this backdrop, Danmarks Nationalbank will focus on climate challenges in a series of publications.

## $CO_2$ concentration in the atmosphere

800,000 BCE to 2019 ACE

The chart shows the number of carbon dioxide molecules per million molecules of dry air.




DANMARKS NATIONALBANK

Source: National Oceanic and Atmospheric Administration (until 1957) and National Aeronautics and Space Administration (from 1958).

# OPPORTUNITIES AND RISKS IN THE RESIDENTIAL SECTOR DURING A GREEN TRANSITION: HOUSE PRICES, ENERGY RENOVATIONS AND RISING ENERGY PRICES

WORKING PAPER

**Alessandro Tang-Andersen Martinello**
Analyses and Models
Financial Stability
Danmarks Nationalbank
alem@nationalbanken.dk

**Niels Framroze Møller***
Climate Task Force
Financial Stability
Danmarks Nationalbank
nfm@nationalbanken.dk

September 9, 2022

**ABSTRACT**

Transitioning to a low carbon economy implies both risks and opportunities in the Danish housing sector, which accounts for one fourth of Denmark's $CO_2$ emissions. We study the heterogeneous impacts on house prices of rising energy prices and energy renovations by combining micro-level data on sales and housing characteristics with data from the official mandatory energy rating reports. We find that higher energy prices will reduce the prices of houses without district heating mainly in rural areas. Most renovations will not increase sales prices enough to cover the costs. Those renovations whose price effect will cover the costs have a lower-than-average impact on $CO_2$ emissions, are cheap, and typically concern houses in and around towns and mid-sized cities and other areas with higher population density and well-developed road networks connected to towns and cities. We conclude that while opportunities for profitable energy renovations are concentrated in these areas, transitional risks are instead associated with peripheral rural areas, where both the exposure to rising energy prices and the risk of financing renovations is highest.

## 1 Introduction

As in other western economies, the residential sector in Denmark is responsible for a large share of society's total greenhouse gas emissions. Since the mid-2000s, the sector has accounted for roughly 25% of total emissions measured in $CO_2$ equivalents.[1] In other words, this sector has substantial potential for climate change mitigation.

A transition to a low-carbon economy will however imply both risks and benefits for homeowners and credit institutions. In this paper, we focus on the role of energy prices and energy renovations, i.e. the renovation of a house with the purpose of improving its energy efficiency.

*We are grateful to The Danish Safety Technology Authority for kindly providing data from condition reports. We thank Tim Thøgersen for scraping and structuring public energy report data. We would also like to thank Thomas Sangill, Thais Lærkholm Jensen, Marcus Ingholt, Martin Oksbjerg, Mette Petry, Ismir Mulalic, Stine Bech, Simone Bonin, Johannes Poeschl, Thomas Harr and Peter Storgaard for comments and suggestions.

[1] Computations based on emission matrices provided by the official Danish statistical authority, Statistics Denmark. Excluding emissions related to car driving the average share over this period remains at around 20%.

The price of energy may rise as a result of changing conditions in global markets that raise demand or cut supply or due to rising carbon- or energy taxes. Through increased heating expenses, particularly for low-efficiency housing, higher retail energy prices can lower the value of homes and thus their collateral value, making financing riskier. Moreover, rising energy prices can compromise debt servicing of some households with low efficiency homes by increasing fixed expenses, thereby adding to the credit risk.

Energy renovations are attractive for households planning to stay in their house for a prolonged period of time when the discounted flow of savings in energy expenditures exceeds the investment cost. They can also be attractive to both homeowners and credit institutions as short-run investments if they increase the sales price more than they cost. However, if the price increase is less than the investment costs, e.g. due to saving flows not being capitalized in the sales price, homeowners have no incentive to improve the energy efficiency of their house if they plan to sell it within few years. In these cases, financing could be associated with increased credit risk.

By exploiting a uniquely rich dataset of housing characteristics and sales between 2014 and 2020,[2] we estimate the effects on house prices of energy renovations and rising energy prices. We focus on the heterogeneity of these effects: First, we exploit flexible causal forests to identify observable drivers of heterogeneity in the effect of energy efficiency and price on house prices (Wager and Athey, 2018; Athey et al., 2019), allowing effects to vary flexibly with characteristics such as geography, year of construction, initial energy efficiency, and overall condition/quality of the house. Second, we exploit the obtained insights to specify more parsimonious and interpretable regression models, which we use to assess statistical significance. We show that ignoring heterogeneity severely biases the sales price effects for the individual households, which for example leads to systematically wrong conclusions as to whether energy renovations are profitable or imply an increase in risk.

Our estimates show that while the magnitude of energy price effects depend primarily on heating source, effects of energy efficiency depend primarily on location. We simulate scenarios on energy price hikes and energy renovations by combining these estimates with information on energy consumption, energy prices, and recommended renovations provided in the official energy reports.[3]

These scenarios show that rising energy prices lower housing prices primarily in rural areas, and for houses not connected to district heating. Specifically, our results imply that a 20% increase in energy prices will reduce house prices in the rural areas by roughly 3-5%.

Moreover, for the majority of households, the effect of renovations on sales prices does not cover investment costs. Insufficient returns are most likely to occur in the urban areas of Copenhagen and Aarhus as well as in some of the more remote and peripherical rural areas (e.g. Lolland, Southern and Western Jutland), while houses in and around smaller towns

[2]For each of the 195,395 sales in our sample, we are able to combine sales prices with information on characteristics such as year of construction, house size, roof type, the number of rooms, floors and bath rooms, plot size, number of buildings, whether a garage or a shed is present, the condition of the house (overall and of the electrical system), as well as geographical location (coordinates). In the scenario analysis, we further exploit detailed data on expected investment amounts and energy savings from the official energy report corresponding to each sale. The sample cut in 2020 (November) results as the official evaluation system underlying the condition reports (proxying quality of a house) was changed subsequently.

[3]These reports are mandatory when selling a house and are performed by independent inspectors. For more information, see https://ens.dk/en/our-responsibilities/energy-labels-buildings (December 9, 2021).

and areas with higher population density, well-developed infrastructure and road networks connected to towns and cities typically enjoy higher coverage rates. Higher coverage ratios are typically associated with cheaper renovations that tend to have a lower effect on $CO_2$ emission than the average recommended renovation.

These results imply that transitional risks are concentrated in some of the more remote and peripherical rural areas. In these areas, house values are typically more sensitive to rising energy prices, and financing energy renovations is riskier if households do not already have sufficient equity to use as a collateral. In contrast, the transitional opportunities, in the form of profitable energy renovations with low absolute costs, are often concentrated in and around towns and in more central areas of Denmark such as Mid and Western Zealand, Funen, and Eastern Jutland.

That the most profitable renovations are those that impact $CO_2$ emissions the least indicates that private incentives may not be sufficient to facilitate climate change mitigation in this context. We show that if home owners financed profitable renovations before selling, $CO_2$ emissions of these houses would have decreased by only 13,000 tonnes per year, or less than 0.02 per cent of total Danish greenhouse emissions. Hence, there may be a scope for policies, e.g. related to tax deductions and the allocation of subsidies for energy renovation among private households, which is currently based on a first-come-first-served principle.

The main contribution of our analysis is to the literature on climate-related financial risks (see e.g. Batten et al. (2016), Krogstrup and Oman (2019), Furukawa et al. (2020), and Bingler and Senni (2022) for discussions and surveys of this literature). As mortgage lenders and banks are linked to the residential sector via their lending, they are exposed to both physical risks and transition risks through debt servicing and collateralization.[4] Most studies have focused on the implications of physical risks (e.g. sea level rise and flooding) for the value of residential properties and mortgage rates. Recent examples are Murfin and Spiegel (2020) and Nguyen et al. (2022) who also survey the related literature, and Mirone and Poeschl (2021) who study the effects of the risk of flooding and sea level rise on the prices of single-family houses in Denmark. In contrast, there has been little focus on financial transition risks coming from the housing sector, in spite of these risks becoming increasingly urgent. Two recent exceptions are Schuetze (2020) who combines energy efficiency data with financial data to show that residential mortages in Germany imply risk exposures towards efficiency standards and carbon taxes, and Ferentinos et al. (2021) who study the effect on house prices of implementing the Minimum Energy Efficiency Standard in England and Wales.

In this paper, we address at least two instances of transition risks. First, by comparing the returns on renovations (the estimated change in sales price) with their costs (from the energy reports), we are the first to shed light on potential risks (or benefits) associated with financing energy efficiency investment. Due to our rich micro-level data and our flexible modeling of heterogeneities in the sales price effects, we are able to draw a particularly granular picture of these risks. Second, our estimates of the impact of effective retail energy prices on sales prices of residential properties allow us to address the risk stemming from

[4]Climate-related financial risks are typically divided into physical risks (e.g. flooding, droughts, wildfires, sea level rise, etc.) and transition risks resulting from the transition to a low-carbon economy (e.g. $CO_2$ taxation, risky new technology, changes in consumer preferences, regulation, etc.). See e.g. Furukawa et al. (2020).

rising wholesale energy prices and taxes. Given the recent dramatic increase in energy prices, shedding light on this risk seems particularly warranted.

Our estimation results also contribute to the literature on energy efficiency and residential sales prices, which goes back to the 1980s.[5] A central theme in this literature is the estimation of the hedonic or implicit price of energy efficiency. Examples, based on data from EPC-type ratings, include studies for Netherlands (Brounen and Kok, 2011), Sweden (Cerin et al., 2014), England (Fuerst et al., 2015), Ireland (Hyland et al., 2013), Denmark (Hansen et al., 2013; Næss-Schmidt et al., 2015) and Germany (Taruttis and Weber, 2022). For the US, analyses have been based on the Energy Star rating program (e.g. Bruegge et al. (2016) and Walls et al. (2017)) and the Home Energy Rebate Program (e.g. Pride et al. (2018)). For Singapore, Deng et al. (2012) study the effect of the Green Mark rating program. The vast majority of studies find that higher energy efficiency increases sales prices of residential housing (i.e. the estimate of the hedonic price of efficiency is positive and significant).

We contribute to this literature by allowing for heterogeneity in the sales price effect of energy efficiency in a systematically data-led way (based on machine learning). Our finding that accounting for heterogeneity is crucial for our main results not only suggests that price effects can be very different across sub-markets. It also casts doubt on the validity of the usual practice of estimating average effects from more aggregate and often pre-specified market segments (e.g. for a country, a state, a national region, or municipality as a whole) compared to averaging sub-market estimates. A further advantage of our study, relative to analyses based on energy ratings (the majority), is that we use data for which efficiency is measured in kWh, which is the underlying variable of energy ratings. This allows us to assess the sales price effect of changes in efficiency even if these changes in kWh are not big enough (but potentially still substantial) to imply a change in the energy label.

The remainder of the paper is organized as follows. In section 2, we explain our empirical approach and the econometric and machine learning models we exploit. section 3 contains a description of the data. In section 4 we present our estimates and in section 5, we combine these estimates with data from the official energy reports in order to simulate scenarios of sales prices. Finally, in section 6 we discuss the policy implications of our results and conclude.

## 2 Econometric framework

Our approach consists of two steps. In the first step, we exploit machine learning methods to identify sources of effect heterogeneity. In the second step, we use these insights to specify interpretable regression models and perform statistical inference. We begin by describing some of the econometric challenges characterizing the analysis of spatio-temporal data. We then describe the double machine learning and the causal forest approach, and finally how we use the results of these approaches to specify the regression models.

[5] Laquatra et al. (2002) provide a survey of the early literature. Surveys of the more recent literature can be found in e.g. Hyland et al. (2013) and Fuerst et al. (2015). The early studies were typically based on data on the past realized energy bills of households. In contrast, later studies have relied on data from energy rating schemes for buildings, as such data have become increasingly available (e.g. the Energy Performance Certificate (EPC) for EU countries and the Energy Star program for the US).

### 2.1 Hedonic price modelling with spatio-temporal data

Our analysis builds on a hedonic price modelling approach (Rosen, 1974). In hedonic models the value of all the utility-bearing characteristics or attributes of a good determine the value of the good. For differentiated goods traded in a market (e.g. houses), prices and attributes (e.g. number of rooms, house size, location, etc.) vary across observations. Empirical hedonic models use this observed variation in the price and attributes of the differentiated good to infer the implicit or hedonic value/price of one or more of these attributes.

In practice, this approach translates into regressing the differentiated good's price on its various attributes and interpreting the estimated coefficients as the estimated hedonic prices.[6] Adjusting for confounding characteristics is crucial for this interpretation to be credible, or equivalently for establishing a causal effect, and we therefore adjust for a wide range of geographical, physical and technical attributes in our models (see appendix B).

We also include variables specifically designed to account for the dependence structure of the data. In a standard cross-sectional regression, it is assumed that observations are generated independently (Hendry and Nielsen, 2007), and violations of this assumption can distort inference (see e.g. Johansen 2006). Cross-sectional observations of realized sales prices that have been recorded over a period of time are most likely not independent due to both spatial and temporal linkages.[7] Spatially, prices corresponding to 'neighboring' sales correlate as they depend on the same factors such as the proximity to environmental amenities, local public goods and services, job opportunities, infrastructure, etc.[8] Temporally, sales prices at any given time are influenced by prices corresponding to recent past sales of houses that are regarded as comparable. This influence is typically referred to as the 'comparable sales effect' (see Isakson 2002 and Small and Steimetz 2012).

We tackle spatio-temporal dependence by conditioning the sales price of house $i$ sold at date $\tau$ on the average square meter price realized in the past year in a neighboring area, $\overline{y}_{i,\tau}$. Similar approaches have been adopted for example by Can and Megbolugbe (1997) and Smith and Wu (2009). For each sold property, we compute the average square meter price of all houses of the same type (i.e. detached single-family homes or townhouses) sold nearby in the past year, starting from the day before the sale date of the specific property.[9]

If we do not condition on the average of prices corresponding to past sales in the neighboring area, $\overline{y}_{i,\tau}$, we are thus likely to have a problem with dependent observations. In addition, not including $\overline{y}_{i,\tau}$ may also imply an omitted variable bias in the estimates of the parameters of interest. In particular, it is possible that past neighborhood sales prices not

[6] The hedonic price may thus be viewed as the marginal willingness to pay for a non-market good. For example, in a house price regression the coefficient corresponding to the number of rooms would be interpreted as the willingness to pay for an extra room.

[7] The general point that independence is too strong an assumption in many applied settings, for example for observations with a spatial or geographical dimension, is well-known in the literature, see e.g. section 2 in Gibbons and Overman (2012).

[8] The spatial dependence problem has led to the use of spatial econometric methods (Anselin (1988)) in hedonic analyses based on house prices. However, in referring to such analyses Thanos et al. (2016) argue that these methods can be inadequate as they are designed to exploit only the spatial, but not the temporal information in the data.

[9] This measure is computed as follows: We start by considering sales within a square area of 100 square meters. If less than five sales occurred in that area in the past year, we consider a square of 1 square kilometer. If less than five sales occurred, we consider a square of 10 kilometers per side. If still less than five sales occurred in the past year, we set the variable as missing.

only influence the pricing behavior connected to a sale of a given house, but that they could also have an effect on energy renovation decisions and hence the energy efficiency of the house. For example, in areas with historically higher prices, where homeowners' equity and hence credit opportunities to finance renovation could be better (Næss-Schmidt et al., 2015).

The general problem of omitted variables has received a lot of attention in the literature of hedonic price modeling due to the spatial nature of the data (see e.g. Kuminoff et al. 2010). Often the regressors of interest (e.g. various non-market urban amenities whose hedonic price one wants to estimate) are spatially varying and co-varying with the omitted variables. In our case, this problem could be relevant not only for estimating the sales price effect of energy efficiency, but also of energy prices as there is geographical variation in this variable also.

The literature has dealt with the problem of omitted spatial variables in different ways (see e.g Kuminoff et al. (2010) and von Graevenitz and Panduro (2015) for discussions). A popular approach for cross-sectional analyses has been to include spatial Fixed Effects (FE), indicating some specified area (e.g. school district, state, part of a city, or municipality). In evaluating this practice and based on detailed empirically calibrated simulations, Kuminoff et al. (2010) show that spatial FEs are often a simple yet very flexible way to deal with correlation between error terms caused by omitted spatial variables that vary across the entities used for the fixed effect and not within.

Because of the large number of observations spread out across the country, we also include municipality dummies as our spatial FEs, thereby accounting for the combined influence from various factors such as municipality tax rate, local public goods and services. Finally, we also cluster the variance-covariance structure according to municipalities.

With $y_i$ representing the natural logarithm of sales prices in sale $i$, we therefore specify our model as

$$y_i = \theta_p(X_i)p_i + \theta_c(X_i)c_i + H_i\gamma + \overline{y}_{i,\tau}\beta + \sum_m \Lambda_m + \varepsilon_i, \tag{1}$$

where $c_i$ and $p_i$ represent the natural logarithm of standardized annual energy consumption in kWh per square meter of heated area at the time of sale, and the natural logarithm of a weighted average of energy prices across all energy types used in a property.[10] The vector $H_i$ represents housing characteristics at the time of sale, including information from condition reports, and $\Lambda_m$ represents fixed effects at the municipality level $m$. We do not place any restriction on functions $\theta_p(X_i)$ and $\theta_c(X_i)$ at this stage, nor on which observable characteristics $X_i$ represents. We explain how we exploit causal forests algorithms to infer the determinants of effect heterogeneity in the next section.

As argued above, including $\overline{y}_{i,\tau}$ contributes to alleviating inferential problems related to dependent observations and omitted variable bias. However, by including $\overline{y}_{i,\tau}$ this regressor will, at any given date, correlate with error terms at previous dates. Hence, an assumption of strictly exogenous regressors, which is often invoked for cross-sectional data, does not apply,

[10]The weights correspond to shares of total standardized consumption as stated in the energy report, see section 3. Standardized consumption must be distinguished from actual consumption and is a measure calculated (by the inspector preparing the energy report) under certain standard or 'average' assumptions about weather conditions, desired temperature, usage of the house, average family size etc. In the present analysis, we allow standardized consumption to include total electricity consumption, as for many of the renovation proposals we consider in section 5, these will also imply saving electricity used for non-heating purposes.

as this requires that all included regressors are uncorrelated with the error term corresponding to each observation. As a result, there is no guarantee that e.g. the OLS estimator (used in the second step of our analysis) will be unbiased under the usual remaining assumptions (see e.g. Hayashi 2000). However, for consistency of OLS we need only that regressors are predetermined (i.e. not endogenous), which requires that they are uncorrelated only with the current error term. That is, if $\varepsilon_{i,\tau}$ denotes the corresponding error term, it holds that $E[\overline{y}_{i,\tau}\varepsilon_{i,\tau}] = 0$. We argue in appendix A that this assumption is reasonable as we construct $\overline{y}_{i,\tau}$ with a sufficient geographical radius and going sufficiently back in time.

### 2.2 Determining relevant dimensions of heterogeneity through causal forests

Our goal is to estimate the effects of energy efficiency and prices on house prices, while flexibly allowing for heterogeneity in these effects across numerous observable characteristics, represented by $\theta_p(X_i)$ and $\theta_c(X_i)$ in equation (1). The standard approach would be to specify subsamples of the data across which we expect the effects to vary and estimate separate regressions for each subsample. This approach is susceptible to multiple testing, which would require specific statistical adjustment, and potentially cherry-picking, with researchers left to navigate a large garden of forking paths (Gelman and Loken, 2014). Moreover, this approach is limited by the pre-specification of subsamples and might miss heterogeneity across unexplored cuts of the data.

Instead, we adopt novel methodological advances exploiting machine learning tools for the estimation of causal effects under unconfoundedness. These approaches provide two crucial advantages. First, by orthogonalizing the main variables of interests via machine learning models, we extract as much information as possible from a given set of controls (Chernozhukov et al., 2017, 2018). Second, by estimating a causal forest model we can identify heterogeneity of effects based on data-driven splits (Wager and Athey, 2018; Athey et al., 2019).

With respect to the model specified in equation (1), we identify sources of heterogeneity for effects of energy consumption and prices using separate models. The two structural models underlying our approach can be written as

$$\begin{aligned} y_i &= \theta^T(X_i) \cdot T_i + q(X_i, W_i) + \varepsilon_i \\ T_i &= f(W_i, X_i) + \eta_i \end{aligned} \tag{2}$$

with $y$ being the outcome of interest, $T$ the main regressor of interest ($p_i$ or $c_i$ in equation (1)), $X$ a set of covariates across which $\theta^T(X)$ can vary, and $W$ eventual additional confounders that ought to be accounted for to ensure unconfoundedness, but which do not determine effect heterogeneity.[11] For each of our models, $W_i$ consists of $\Lambda_m$ and $\overline{y}_{i,\tau}$, while $X_i$ contains any characteristic in $H_i$ plus either $c_i$ or $p_i$, whichever of these variables do not already enter the model as the main variable of interest $T_i$.

We use flexible machine learning models to provide estimates of $\hat{q}(X, W)$ and $\hat{f}(X, W)$. This first-stage orthogonalization allows us to estimate effects of $T$ in a second stage and can be then thought as a generalization of the Frisch–Waugh–Lovell theorem for linear regressions. We perform model training and successive orthogonalizations in separate subsamples. In practice, we split our sample into two random subsets, and fit a machine

[11] In order to facilitate replication of our approach, this paper adopts as much as possible the notation used in Battocchi et al. (2021), which is the python package used for estimation.

learning model in each subset. We then use the model fitted on one subset to orthogonalize the other subset, and vice versa. Due to their high performance in tabular data sets, we use gradient-boosted trees as first-stage models for both $T$ and $Y$ (Ke et al., 2017).

We identify sources of effect heterogeneity by estimating the models through causal forests. Building on Athey and Imbens (2016), Wager and Athey (2018) propose to estimate $\theta(X)$ through an ensemble of causal trees. While standard tree-based models operate by iteratively splitting a sample into subsamples (leaves) based on observable characteristics such that differences in mean outcomes are minimized within a leaf and maximized across leaves, causal trees do the same for differences in effects, estimated via sample splitting. While half of the data is used to estimate the tree structure, the other half is used to estimate effects within each leaf. In a causal forest, the individual estimate $\theta(X_i)$ will then be the average of the effects estimated within each leaf in which the observations are assigned across all trees (Athey and Imbens, 2019).

We use the causal forest approach to provide insights on heterogeneities existing in the data across combinations of observable characteristics $X$. Given these insights, we then proceed to estimate these effects based on a standard regression model and OLS. This allows us to interpret our results clearly and provide statistical inference by taking into account spatial dependence across error terms.

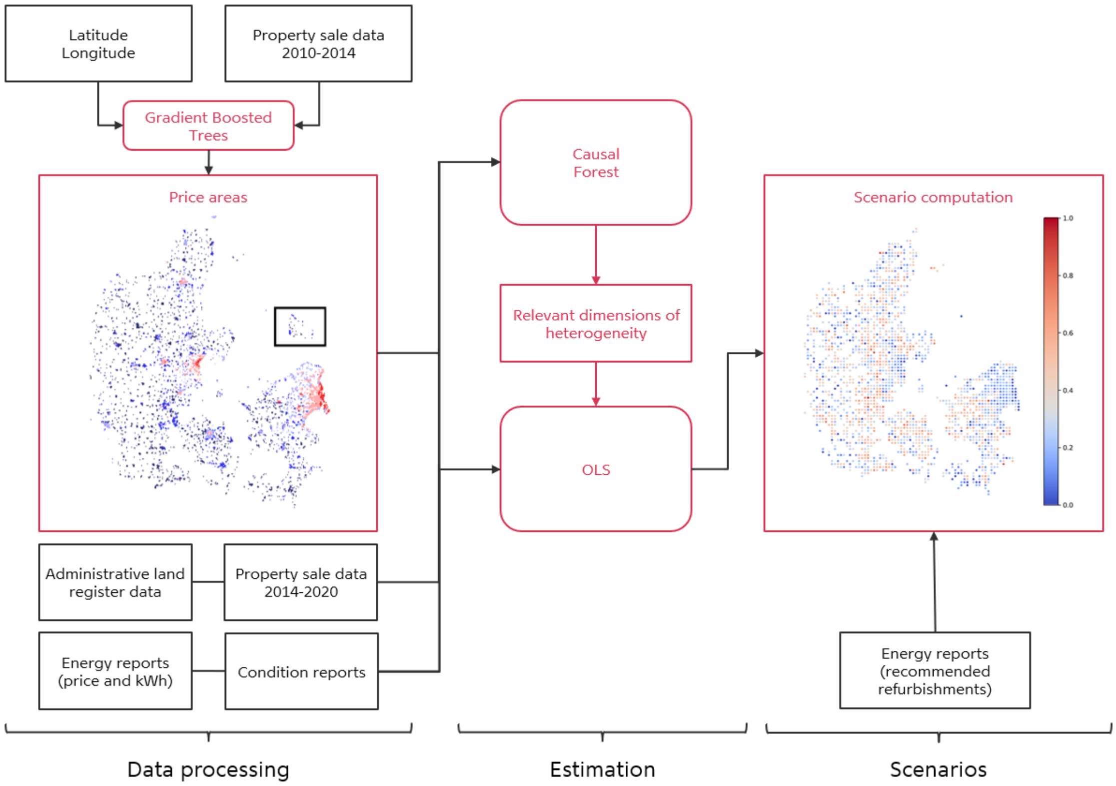
### 2.3 Inference based on linear regression

Once the ML algorithm has identified the heterogeneities, we can partition the data into subsamples and use linear regression to estimate potentially different effects from energy efficiency and prices for each sub-sample. As we have argued, the parameters in this regression can be estimated consistently by OLS as long as care is taken in modeling the dependence structure of these spatial-temporal data. We do so by conditioning on both average neighborhood sales prices from the past and municipality-fixed effects.

An obvious alternative approach would be to use the results from the ML analysis to fit one regression equation with the coefficients on standardized energy consumption and energy prices allowed to vary according to the identified heterogeneity. In other words, a model within which these two key variables ($T$) are allowed to *interact* with other explanatory variables ($X$). However, as pointed out in Balli and Sorensen (2013), because variables in $X$ can be correlated with other explanatory variables in the model, the interpretation of the estimates corresponding to the main interactive terms (symbolically denoted $X \times T$) is not obvious. Specifically, controlling only with linear (non-interactive) terms of these other variables is not enough to avoid that $X \times T$ picks up the effects from the interaction of some of these with those in $T$. While, as the authors suggest, one could orthogonalize regressors (or alternatively include *all* relevant interacting terms), we take the simpler approach of splitting the data into sub-samples, since we can afford this extra flexibility given the large number of observations in our sample.

Finally, we use the coefficients estimated through these regressions to extrapolate counterfactual scenarios and compute the effect on house prices of implementing the recommended renovations as specified in the energy reports, or of increasing energy prices. Figure 1 visualizes the steps of our approach, and the raw data we use in each of them.

**Figure 1**
**Visualization of estimation approach**




NOTES: The figure shows the analytical process of this paper. After processing the data (see section 3), we use a causal forest model to detect relevant dimensions of heterogeneity. Armed with this knowledge, we estimate effects through OLS and use those estimates to finally extrapolate scenarios informed on data on energy prices and recommended renovations contained in the energy reports. Steps highlighted in red are model-based, while black blocks represent raw data.

## 3 Data

We combine data from multiple administrative registers into a unique dataset covering over 195,000 sales of owner-occupied single-family homes between 2014 and 2020. The main building block of our dataset is third-party reported property sale data, which contain information on sales price, time of sale, and the municipality in which the property is located. We join this information with data from the administrative land register, containing detailed information on a wide range of structural characteristics of the property sold, from building and plot size through its location in the form of geographical coordinates and the type of roof. A detailed description of the variables used in the analysis appears in appendix B.

We further enrich this data with information from the energy and condition reports available for the properties in our sample. Both energy and condition reports are prepared by independent inspectors. During our sample period energy reports are mandatory in connection with a sale.[12] Condition reports are prepared for the vast majority of sales. An advantage of these data is therefore that we do not have self-selection into providing

[12]In 1997 energy rating of housing became mandatory in connection with sales, and since 2010, real estate agents must state the label in sales advertisements.

an energy report (see e.g. Brounen and Kok 2011, section 5.2, and Hyland et al. 2013). Energy reports contain detailed data about the expected energy consumption per square meter calculated by the inspector and energy costs/prices for the given property. Therefore, standardized energy consumption and effective retail energy prices (i.e. the total price per kWh including fixed costs, that is paid by the end user) are available for all houses in our sample.

There is substantial variation in effective retail energy prices across observational units. This variation persists even across households that use the same and only one form of energy. For example, for the most widespread heating source in Denmark, district heating, prices differ markedly between different areas, since these have different utilities supplying under different circumstances.

There are two types of condition reports for a property. The first contains an assessment of the general condition of the building(s). The second concerns exclusively the electrical installations. The condition reports list the number of minor to severe damages, defects, and omissions, and are a good proxy for the quality of the property at the time of sale (see appendix B).[13] Overall, these data provide a uniquely rich overview of housing characteristics, and allow us to control for a range of possible confounders when estimating the effect of energy efficiency and energy prices on house prices. Out of sample, the data we collect are able to explain up to 80% of variance in square meter prices.

As emphasized in section 2, due to spatio-temporal correlation between observations it is important to include geographical variables. Hence, we include both municipality fixed effects and the average square meter price realized in the past year in a neighboring area, $\overline{p}_{i,\tau}$. However, while the latter regressor provides a good time-varying proxy for local amenities, it might be particularly noisy in times and areas with static housing markets and few housing transactions.

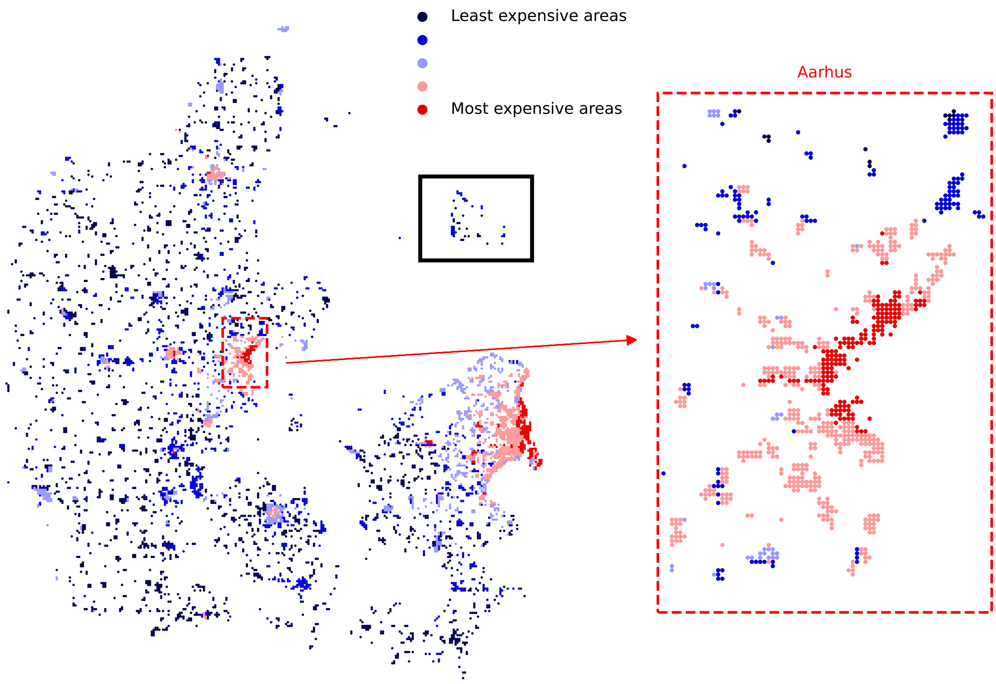
Therefore, as a third geographic indicator we construct a time-constant proxy for amenities and otherwise unobserved local drivers of price differences by mapping geographic coordinates in five price areas. We construct these areas by extracting a random sample of house sales among those we do not consider in our analysis (i.e. houses sold before 2014). We then split these sales in five quintile groups according to their square meter price, normalized within each calendar year, and train a gradient boosted trees model to assign the probability that a house will belong to each quintile group given only its geographic coordinates (Adolfsen et al., 2022). This approach identifies areas with generally high price levels across the country and within municipalities. In equation (1) these areas enter as five probability vectors as part of housing characteristics $H_i$.

Our estimated price areas appear in figure 2, which shows that higher price areas tend to cluster in town centers or around popular amenities, with lower price areas representing remote countryside locations and suburban areas neighboring highways and factories. The figure also shows that this approach identifies substantial variation in price areas even within a single municipality. For example, all five types of areas are represented within Aarhus Municipality, the second most populous city in the country. This approach enables us not only to account for granular geographical variation within municipality and zip codes, but

[13]Data for the general condition of a house were also included in Næss-Schmidt et al. (2015), but not in Hansen et al. (2013).

also to parsimoniously represent this variation in only five dimensions, allowing us to use price areas as interpretable determinants of effect heterogeneity.

**Figure 2**
**Estimated price areas in Denmark**




NOTES: The picture maps the estimated price areas in Denmark, further zooming into the second largest city in the country, Aarhus, which contains all five price areas within its territory. The five areas are estimated using a random sample of about 66,000 house sales between 2010 and 2014, which are not included in our analysis sample. We estimate these areas in two steps. First, we divide sales into five equally sized groups sorted by the realized square meter price, normalized to have mean zero and a standard deviation of 1 within each calendar year. Second, we train a gradient-boosted trees model to predict, given the geographic coordinates of the house, in which of the five groups the house is most likely to belong. This procedure returns a mapping of coordinates to price areas both across and within municipalities.

We remove from our sample a limited number of observations (less than 0.1 per cent) having unrealistically high energy prices as a result of erroneous recordings. We also remove seven observations for which the standardized consumption, the heated area, or the energy price is zero or negative. This minimal sample selection leaves us with a sample of 195,395 sales for our analysis.

## 4 Effects of energy efficiency and prices on house prices

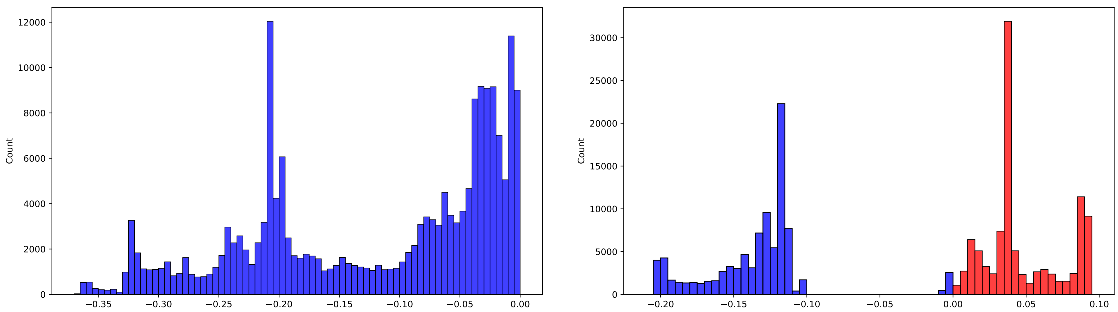
This section presents our estimates of the heterogeneous effects of energy price and standardized energy consumption (reciprocal efficiency) on residential sales prices. We proceed in two steps. First, we identify drivers of effect heterogeneity through our causal forest models. Second, we exploit the insights generated by the causal forest model to design and estimate separate regressions by OLS, one for each combination of control variables that is relevant for effect heterogeneity. Through this process, we retrieve heterogeneous effect estimates on interpretable subgroups, while both allowing for potentially any combination of control variables to affect our coefficients of interest and performing valid statistical inference.

In order to document the importance of taking heterogeneity in effects into account, in this and the next section, we compare our estimates with estimates based on a model in which the effects of energy efficiency and prices on sales prices do not vary with any characteristic. The results of this estimation are found in appendix C.

### 4.1 Identifying effect heterogeneity

The causal forest model allows the estimation of Conditional Average Treatment Effects (CATEs) for every observation in the sample. Figure 3 shows the distributions of these estimated effects for both standardized energy consumption and energy prices, revealing substantial variation around average effects. For standardized energy consumption, estimated elasticities range approximately between -0.4 and 0, compared to an average estimated effect of -0.12 (see appendix C)

**Figure 3**
**Distributions of Conditional Average Treatment Effects (CATEs) estimated by causal forests**



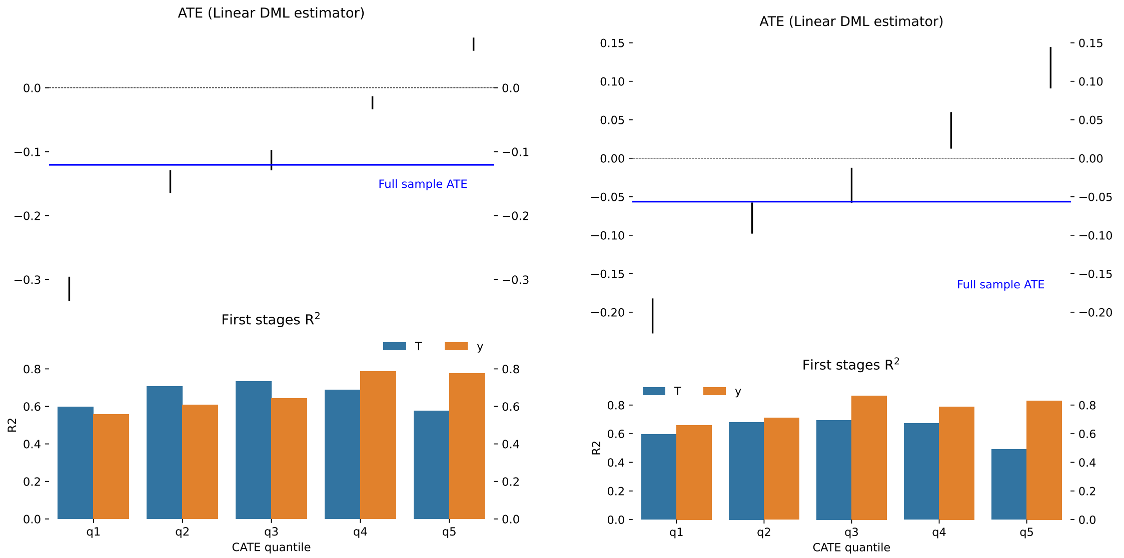

**(a) Standardized energy consumption** **(b) Energy price**

NOTES: The figures show the distributions of Conditional Average Treatment Effects (CATEs) estimated by separate causal forest models for standardized energy consumption and energy price. Blue bars indicate negative effects, red bars positive effects. For each model, the effect of the variable of interest is allowed to vary by housing characteristics (see tables 5 and 4 in the appendix), with the exception of municipality fixed effects (too granular) and neighborhood price per $m^2$ ($\overline{y}_{i,\tau}$, time-varying).

For energy prices, the heterogeneity is even more staggering. The causal forest estimates, that for no properties effects are approximately around -0.056, the estimated average effect based on the model with homogenous effects. Rather, the distribution is separated in two, with a group of properties characterized by large negative effects and another group characterized by small positive effects.

We follow the approach of Athey and Wager (2019) and show in two ways that these estimates capture real heterogeneity in the data. In the first approach, we divide the distributions plotted in figure 3 in five equally-sized groups based on quintiles. We then estimate separate double machine learning models (Chernozhukov et al., 2018) and estimate the coefficient of interest within each of these groups. If the causal forest correctly captures heterogeneity in effects, these coefficients should be monotonically increasing as we move from the lowest to the highest quintile of the estimated distribution.

**Figure 4**
**Quantile test: The causal forest models capture heterogeneity of effects in the data**




**(a) Standardized energy consumption** **(b) Energy price**

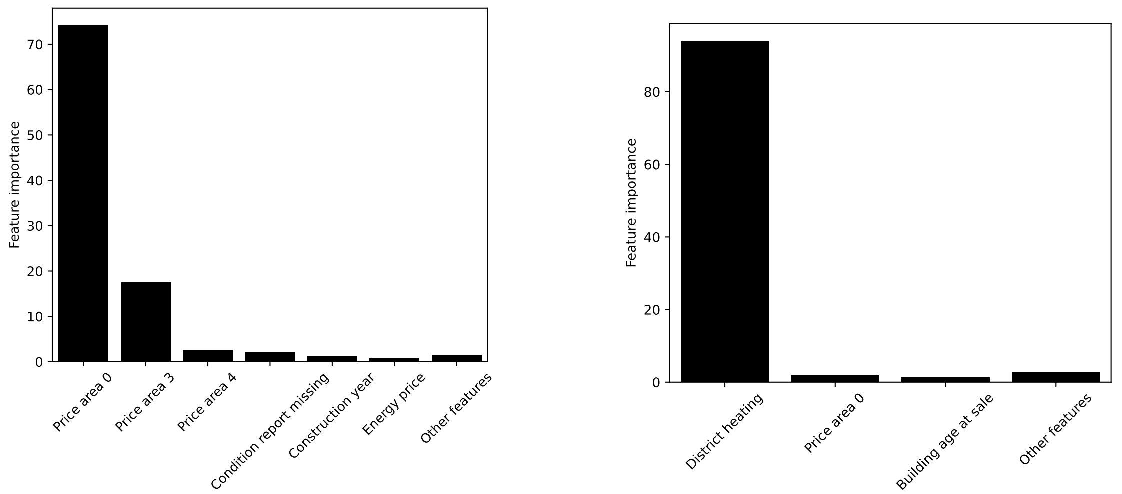
NOTES: The top panels of the figure plot the effect of interest within each CATE quantile estimated through separate double machine learning models (Chernozhukov et al., 2018). Bars indicate 95% confidence intervals, with observations assumed to be independent. The bottom panels of the graph show for each quintile the estimated out-of-sample $R^2$ for both the regressand and the main regressor of interest.

The top panels of figure 4 show that coefficients are indeed monotonically increasing for both the effect of standardized energy consumption and energy price. Figure 4 also shows that we can explain up to 80% of the out-of-sample variation in real estate prices and standardized energy consumption, and up to 75% of the out-of-sample variation in energy prices, confirming that our control variables account for the majority of what determines house prices. The top quantile of estimated CATEs of energy prices, for which we estimate positive effects on housing price, is also the group for which we can explain the least variation in energy prices given our controls. This pattern points to us not being able to adequately alleviate omitted variable bias in this group.

In the second approach, we compute the test for effect heterogeneity proposed by Athey and Wager (2019). The test builds on the estimated CATEs and the first-stage model residuals and reveals whether the causal forest model is well calibrated. The results of the test appear in table 7 in the appendix and show that our causal forest models are adequately, yet not perfectly calibrated. Therefore, we use these models primarily to provide insights on the drivers of effect heterogeneity and do not use the estimated CATEs directly.

To understand the drivers of effect heterogeneity, we interpret the causal forest models by computing feature importances, i.e. the normalized number of times a single regressor has been used by the model to identify a leaf. Despite being silent about the sign of a regressor's influence on the estimates, this metric provides a first approximation of how much a specific regressor matters for effect heterogeneity.

**Figure 5**
**Feature importance**




**(a) Standardized energy consumption** **(b) Energy price**

NOTES: The figure plots feature importances in the second-stage causal forest model. These scores reflect the number of times any single regressor is actively used by causal forest trees to split a node and create a new branch or leaf in the tree. As such, the figure does not represent a variance decomposition, but rather how often the model uses a specific regressor to separate across groups of houses for which energy consumption has different effects.

We find that few observable characteristics play a substantial role for effect heterogeneity. Figure 5 shows that the house price elasticity with respect to standardized consumption primarily depends on the price area and to a lesser extent on whether a condition report was present at the sale or not. For the energy price effect, the decisive factor is whether the heating source is district heating or not. The price area to which the property belongs plays a minor role.

This interpretation is further supported by figure 10. We train single decision trees on the estimated CATEs as regressand and using $X$ as regressor. These trees split groups according to price areas and missing condition reports in the case of standardized energy consumption, and according to district heating and price areas in the case of energy prices. In the following, we therefore chose to focus on price area, the existence of a condition report, and district heating as the characteristics to describe the heterogeneity in the price elasticities of standardized consumption and energy price.

### 4.2 OLS estimation of the heterogenous effects

We estimate the sales price elasticity with respect to standardized consumption through ten separate regressions corresponding to the ten combinations of price area and presence/absence of the condition report, with differences in standardized consumption across houses reflecting a mix of e.g. differences in insulation and the efficiency of the heating source. Likewise, we estimate the elasticity with respect to energy price for each of the ten combinations of price area and heating source being/not being district heating. In all regressions, we include all remaining controls, and cluster the variance-co-variance matrix by municipality.

The estimated heterogeneous price elasticities appear in table 1. For all price areas the estimated elasticities with respect to standardized consumption are negative and significant except for the most expensive area. These findings hold regardless of whether the condition report exists or not. However, for all areas the estimated effect is larger when the condition report is absent. This pattern occurs because without a condition report we cannot control for the condition or quality of houses. As a result, the effect from standardized consumption will be overestimated, since a higher standardized consumption (lower efficiency) is typically associated with a poorer general condition of a house.[14] Finally, the elasticity with respect to standardized consumption is numerically lower the more expensive the area is.

The estimated effect of energy price differs dramatically across heating source. For houses that do not use district heating as their primary energy source, energy prices reduce house prices in the less expensive areas of the country, as in these areas energy expenditures constitute a larger share of housing costs. For houses with access to district heating, our results seem to indicate that energy price increases house prices. We attribute this result to the lack of exogenous variation in energy prices for this subsample. In Denmark, for many areas with district heating there is an obligation to connect to the district heating net: Energy prices are often fixed within geographical areas, and we do not have sufficient insights on how they are determined and on whether they might endogenously depend on house values.

**Table 1**
**Estimated heterogenous price elasticities**

| | Standardized energy consumption | | Energy price | |
|---|---|---|---|---|
| Price area | No condition report | Condition report | No district heating | District heating |
| 0 (least expensive) | -0.526<br>(-16.576) | -0.347<br>(-29.932) | -0.267<br>(-17.088) | -0.003<br>(-0.087) |
| 1 | -0.423<br>(-9.616) | -0.204<br>(-12.496) | -0.159<br>(-6.216) | 0.115<br>(2.720) |
| 2 | -0.328<br>(-8.131) | -0.118<br>(-5.610) | -0.094<br>(-3.865) | 0.157<br>(3.045) |
| 3 | -0.157<br>(-5.638) | -0.061<br>(-3.751) | -0.003<br>(-0.144) | 0.211<br>(5.292) |
| 4 (most expensive) | -0.039<br>(-0.975) | -0.004<br>(-0.268) | 0.090<br>(2.725) | 0.231<br>(2.994) |

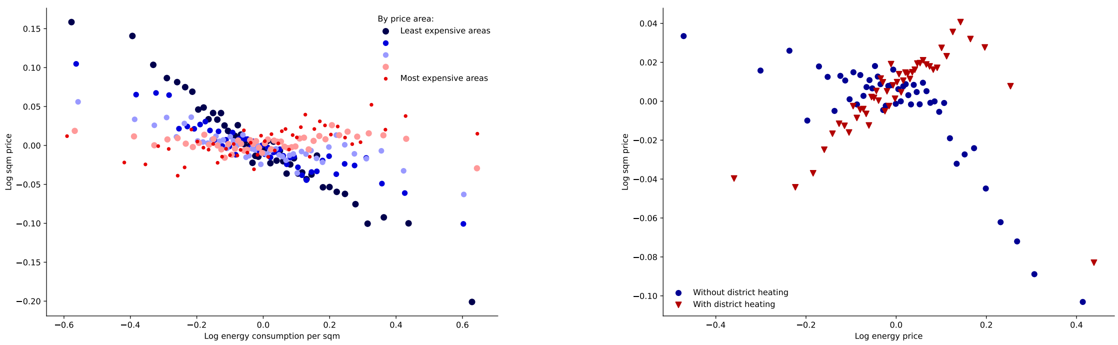
The table reports the estimated elasticities of sales prices with respect to standardized energy consumption and energy price. The estimation is based on linear regression models estimated with OLS on separate subsamples of the data. The t-statistics (in parentheses) are based on standard errors clustered at the municipality level.

Overall, there is a large difference between the estimated coefficients for energy price and those for standardized energy consumption. If these two variables only influence sales prices via the variable part of the energy expenditure, i.e. via their product, their coefficients

[14]This finding fits with the fact that Næss-Schmidt et al. (2015), whose data include condition reports, find a lower effect from energy label (rating) than that found in Hansen et al. (2013) that do not include this information. In the latter study, the authors do not have information from condition reports, but are aware of the bias it may introduce.

should be identical. One interpretation of the difference in coefficients can be if buyers prefer higher energy efficiency (lower standardized consumption) - not only because it implies lower energy costs in monetary terms but also because it implies lower emissions. In such a case, the effect from standardized consumption will be greater than the energy price effect. A more irrational behavior could also explain the difference. For example, due to the emphasis on the energy rating in connection with a sale, the buyer could place relatively more weight on energy efficiency than on the effective cost per kWh.

**Figure 6**
**Partial regression plots**




**(a) Standardized energy consumption** **(b) Energy price**

NOTES: The figure shows a binned residualized scatterplot of the relationship between log house prices per square meter and key variables of interest, divided by price areas (left) and heating source (right). Each plot residualizes the main regressor and the regressand using the estimated first-stage gradient-boosted trees models. For each price area we then bin the residuals in 50 approximately equally sized bins according to the regressor residuals, such that each bin represents approximately 2% of the data within each group. Finally, we plot the average of the regressand and regressor residuals against each other for each bin in a scatterplot, where the size of each point is proportional to the number of observations it represents.

Figure 6 supports the choice of a log-log specification for the separate regressions. The figure shows binned scatterplots of the residuals of log energy price and log standardized energy consumption against log square meter house prices, when adjusting for all remaining controls. We split each plot by the most relevant dimension for effect heterogeneity.

The left panel not only shows that the conditional relationship between house prices and energy consumption is linear in log space, but also how the slope of the relationship changes across price areas. The relationship is strongest for rural, cheaper areas, and approximately zero for expensive urban areas.

The log-linear form implies that a given absolute decrease in standardized consumption will imply a higher absolute price increase, the lower the initial standardized consumption is. Hence, the marginal willingness to pay for extra efficiency increases with efficiency. This pattern could reflect the limited supply (relative to demand) of houses with very high efficiency and/or that sub markets for high-efficiency houses are in general very different from submarkets for houses of lower efficiency.

The right panel shows that the relationship between house and energy prices also changes starkly depending on whether the house uses district heating or not. For houses using alternative heating sources, the relationship is negative and roughly linear. For other houses, the relationship is primarily positive, and negative for the highest percentiles of conditional energy price. Figure 11 in the appendix further splits these partial regression plots by price areas. Mirroring the results in table 1, the figure reveals that the positive association between energy and prices of houses without district heating is not present in rural areas and is relatively noisy in other areas.

District heating is primarily present in urban areas, and its pricing not only varies geographically, but is also potentially endogenous. We acknowledge that our analytical approach is not able to tackle such endogeneity. Therefore, when discussing policy implications of rising energy prices, we will focus on energy prices for houses that do not use district heating as a heating source.

## 5 Scenarios for energy renovations and rising energy price

This section combines the results presented in section 4 with information from the energy report on both energy prices, energy savings, and investment costs corresponding to the proposed energy renovations. Through this combination, we investigate whether energy renovations can be expected to increase house prices enough to cover the initial investment costs. Moreover, we assess how much it will cost society and how much total $CO_2$ emissions can be reduced when households renovate. Second, we assess the potential impact on house prices of future increases in energy prices and/or taxes.

### 5.1 The transitional risks and opportunities of energy renovations

We focus on renovation proposals that in the energy report are classified as being profitable. Reports classify an investment as profitable if the implied energy savings from renovation will cover the initial investment cost before the installed device needs to be replaced. For these proposals, the report contains an estimate of the costs of the proposed investment in addition to its implied savings.[15]

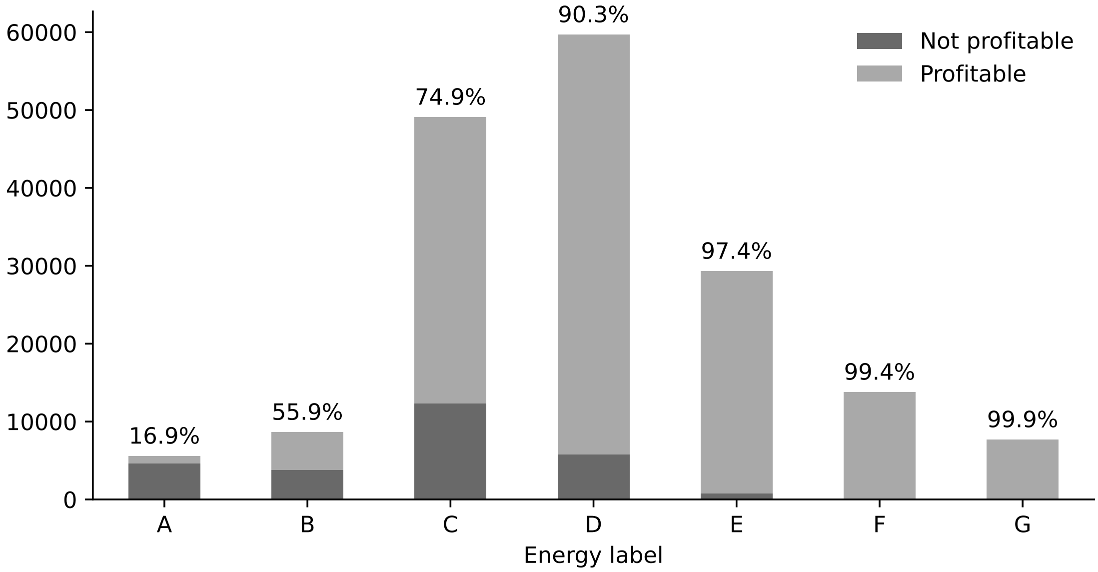
In the renovation scenario, we assume that all households carry out all profitable investments proposed and only these. On the one hand, this assumption may be viewed as an upper limit for the extent of renovation and the resulting credit demand. For example, homeowners could regard the individual proposals more as alternatives, some of which are relatively more profitable than others. On the other hand, households may prefer energy renovation in connection with other reconstructions or extensions of the house. In that case, the household's total investment costs and credit demand could be substantially bigger. However, our results do not provide information about the increase in house prices that could result from combined renovations (see footnote 15).

For each level of energy efficiency measured by the energy rating, figure 7 shows the total number of houses sold and the share with profitable proposals. As expected, the figure shows that it is mainly among households with medium to low efficiency (energy labels C through G) for which profitable investments are possible.

We compute how much the sales price will increase if a given household carries out all profitable renovations as the difference between the regression function that applies for

[15]The report also contains proposals of energy renovations that should be combined with other reconstructions or extensions of the house. The expected costs of these are, however, more uncertain and therefore not stated in the report. In addition, some proposals are classified as 'recommended', but as this is optional for the inspector to fill in, it is often left empty.

**Figure 7**
**Distribution of profitable proposals by energy rating/label**




NOTES: The figure shows, for each energy rating/label, the number of houses in our sample split by whether the energy efficiency renovation described in the report is deemed profitable or not by the inspector. A is the label for the highest energy efficiency levels (covers labels A2020, A2015 and A2010 of the official scale. See www.ens.dk). Label G indicates the lowest efficiency. On the top of each bar the figure reports the share of profitable renovations in the subsample. The category 'Not profitable' corresponds to either no proposals or the aforementioned proposals that should be combined with other renovations/reconstructions of the house.

this household evaluated in the present level of standardized consumption and evaluated in the hypothetical level after all profitable renovations have been carried out. We only use the estimates for which the condition report is available. Hence, the estimated regression functions that apply for the individual households correspond to those underlying table 1, third column.

Table 2 compares estimated house price increases and investment costs. The upper part of the table presents results based on estimation that takes heterogeneity into account. The first column shows the estimated average price increase resulting from profitable renovations. The larger OLS results are our preferred estimates. Causal forest models are not perfectly calibrated, and therefore we interpret results based on those estimates as lower bounds.

The last three columns compare the price increase resulting from renovation to the investment cost. We focus on the price return ratio on investment, which represents the estimated house price increase divided by the invested amount. The ratio thus reflects the house price return in kroner per 100 kroner invested. Overall, the table shows that energy renovations do not typically increase house prices enough to compensate for the investment costs. While there are some investments that have a very high return ratio due to low investment costs, the median return ratio is well below 1 no matter the model chosen once effect heterogeneity is accounted for. For most houses, the investment cost of energy renovations would not be fully reflected in the sales price.

**Table 2**
**Renovation scenario: Main results for all of Denmark**

| | Average renovation effect (thousands kr.) | Median price return ratio on investment | Mean price return ratio on investment | Percentage with return ratio > 100% |
|---|---|---|---|---|
| | *Heterogeneous effects* | | | |
| Causal forest | 44.23 | 31.28 | 69.95 | 15.09 |
| OLS | 72.31 | 62.90 | 120.19 | 30.06 |
| | *Homogeneous effects* | | | |
| Double machine learning | 73.84 | 59.21 | 106.22 | 28.86 |
| OLS | 149.28 | 117.68 | 211.10 | 56.86 |

NOTES: The first row computes renovation effects through the CATE estimates presented in the left panel of figure 3. The second row computes renovation effects through the estimates presented in table 1, where coefficients that are not statistically significant at the 1% level have been set to zero. The last two rows compute renovation effects through the estimates presented in the last two rows of table 6.

The last two rows of the table, which extrapolate effects of renovations from the benchmark models (with homogenous efffects) presented in table 6 in appendix C, highlight the importance of accounting for heterogeneity. The differences in results relative to the first two rows are striking: First, the estimated renovation effects (price increases) are much higher when assuming homogeneity reflecting highly overestimated average effects. Second, the main conclusions change. By ignoring effect heterogeneity, it would appear that the majority of renovations have return ratios higher than 100% when using OLS results.

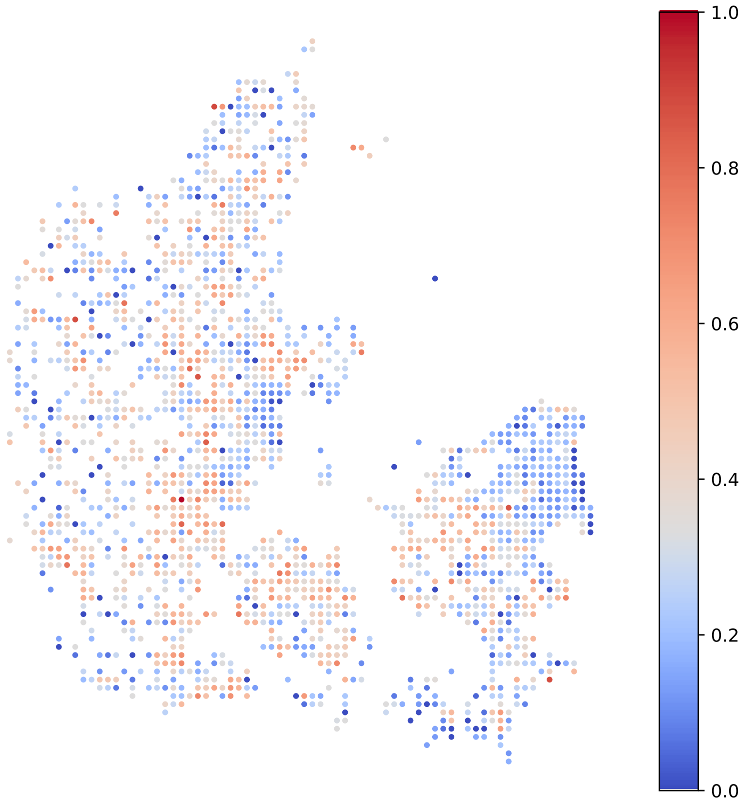
To sum up, table 2 shows that for the great majority of the houses in the sample the returns (price increase) do not cover investment costs. As we only focus on profitable renovations in the long run, this result implies that the future flow of energy savings is not fully capitalized into sales prices.[16] Taking heterogeneity into account is crucial for reaching these conclusions. We discuss the implications of these results further in section 6.

Nonetheless, the aggregate results from table 2 hide substantial geographical variation. Figure 8 highlights the geographical variation by dividing the map of Denmark into 4x4 kilometer grids. For each grid, we compute the share for which the price increase is greater than the investment cost. Deep red indicates that the share is close to 1, while the dark blue indicates that it is close to 0.

In the urban and suburban areas of Copenhagen and Aarhus, the price increase rarely covers the investment cost. This result also holds for low-price rural areas, such as rural Lolland, and parts of Western and Southern Jutland, despite houses in these areas having lower energy efficiency on average. Hence, despite rural areas having the largest percentage price increases, measured in absolute terms (kroner), these increases rarely cover the invested amount. In rural areas, this result is mainly due to very low sales price levels and to a less extent higher investment costs. In urban areas, the primary reason is that house prices are generally insensitive to energy efficiency. One explanation for this insensitivity could be that energy expenditures in general constitute a small share of the overall housing costs in urban areas.

[16]This result is in line with the findings in Næss-Schmidt et al. (2015).

**Figure 8**
**Returns on recommended renovations are lowest in cities and low price rural areas**

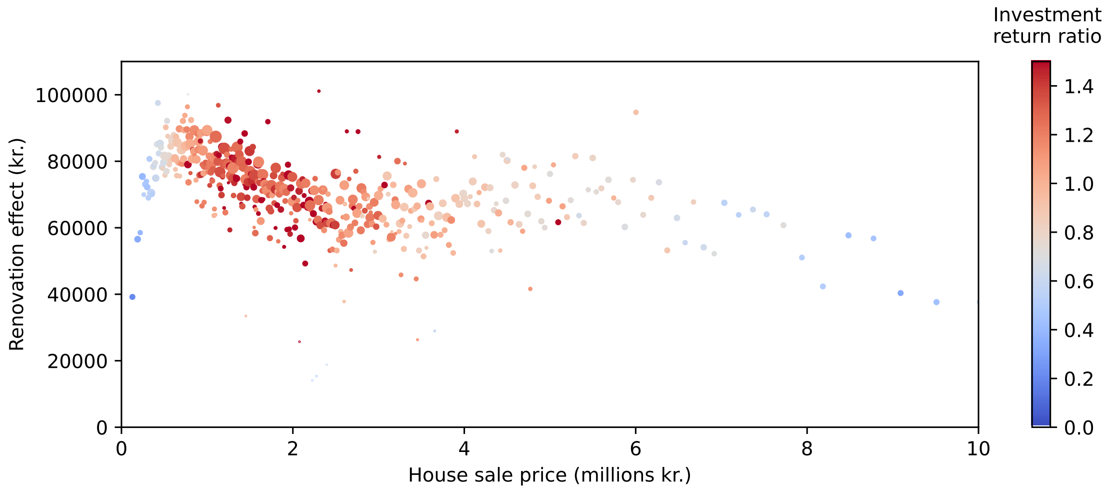



NOTES: The figure divides Denmark in squares of 4x4 kilometers. Within each square, the color scale indicates the share of houses whose price increase after a renovation would be higher than the expected investment cost in the report. The renovation effects of house prices are computed through the estimates presented in table 1, where coefficients that are not statistically significant at the 1% level have been set to zero. Squares with less than five sales occurring in our sample are not shown.

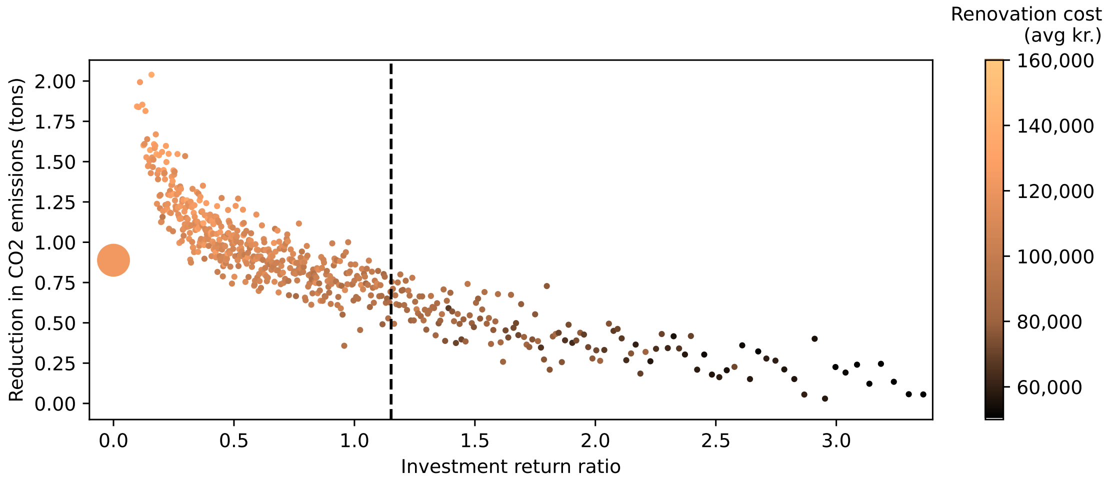
Figure 8 shows that although for the majority of houses in the suburban/urban areas of the biggest cities and some low-price rural areas energy renovations do not pay off in the form of an increase in sales price, there exist a sweet spot in which return rates are considerably higher and often above 1 (100%). These areas are large parts of Eastern Jutland, Funen and Mid and Western Zealand, but also specific areas in and around certain towns (e.g. Vejle, Ribe, Silkeborg, Roskilde, and Nykøbing Falster). The top panel of figure 9 shows that the sales prices of these houses are typically within a range of roughly kr. 1-2.5 mio. For more expensive houses with similar or slightly lower absolute price increases (renovation effects), the return ratio decreases.

The renovations associated with houses in this sweet spot are, however, not those that make the largest dent in the housing sector emissions. The bottom panel of figure 9 shows that the return ratio of energy renovations in terms of house prices is negatively correlated with the return in terms of $CO_2$ emission reductions. Renovations, the costs of which are covered by the price increase (return ratio $\geq$ 100%) and which are therefore less risky in the short run for an investor, on average reduce a relatively small amount of emissions. On the

**Figure 9**
**Investment return ratios are related to house prices, reductions in $CO_2$ emissions, and investment costs**




**(a) Houses the price of which is between kr. 1 and 2.5 million tend to have the highest investment return ratios.**




**(b) Renovations with high investment return ratios tend to be cheap, and to impact $CO_2$ emissions the least.**

NOTES: The figure is based on estimates from the heterogeneous linear regression models (OLS). For each observation in our sample, we combine our results with data from both the sale and the energy report to compute the effect of the recommended renovation on house prices and its investment return ratio (effect on house price over investment costs). The effect for houses for which the estimated renovation effect is insignificant is set to zero. These measures are then compared to the expected renovation cost and reduction in $CO_2$ emissions. In each graph, we group the full sample in approximately 50 equally-sized bins given the value of the x-variable. Groups for which the x-value is the same are pooled together. We then plot averages within each bin as a scatterplot, indicating with market size the relative size of the bin.

one hand, this relationship shows that the private economic incentive for reducing $CO_2$ is small. On the other hand, houses in the sweet spot can be renovated at a relatively low cost with a positive return on the investment and still contribute to $CO_2$ reductions.

We confirm this interpretation through a simulation exercise, where we assess the costs and benefits of renovating subsets of the houses in our sample before a sale (table 3). We

**Table 3**
**Simulations of aggregate impacts on house values and $CO_2$ emissions of selected renovations**

| | Renovate top 25% of houses in terms of house price return ratios | Invest the same amount to reduce as much $CO_2$ emissions as possible |
|---|---|---|
| Total investment cost, kr. million | 2403 | 2403 |
| Sum of house price effects (causal forest), kr. million | 2668 | 1100 |
| Sum of house price effects (OLS), kr. million | 4749 | 1714 |
| $CO_2$ saved, tonnes | 13597 | 66153 |
| - CPH-NYC flights | 20601 | 100232 |
| - Carbon offset cost, kr. million | 8 | 39 |

NOTES: The first column computes aggregate figures given renovating before a sale the 25% of houses for which the ratio between the price effect of a renovation and its cost is highest. These renovations cover over 36,000 houses. In the second scenario, we simulate renovating as many houses as possible given the same invested amount, starting from houses for which emissions are reduced the most for each invested krone. This scenario represents the maximum achievable reduction in emissions for a given total investment. We price each tonne of $CO_2$ according to the EU Emission Trading System futures price. At the time of performing this calculation (December 6 2021), the EUA futures price was at a record high of 81.25€ per tonne.

consider two strategies. In the first, we select 25% of the houses to renovate based on their return ratio. These renovations are incentive-compatible, in the sense that the costs of renovations will be covered by the short-run increase in sales prices: Investing in such renovations before a sale will return a profit, such that owners have the incentive to renovate both in the short and in the long run. While performing these renovations will cost over kr. 2 billion, these costs will be covered by the increase in sales prices even considering the lower bound estimate provided by the causal forest model. These incentive-compatible renovations would nonetheless save over 13,000 tonnes of $CO_2$ emissions yearly. These results point to an unrealized potential of reducing emissions in the residential sector.

We compare this strategy to one in which the available funding for investments is the same, but the houses to renovate before selling are selected by the amount of $CO_2$ emissions reduced per invested krone, as proposed by Lang and Lanz (2022).[17] These renovations would not be incentive-compatible in the short run. Given an expenditure of kr. 2.4 billion, the total increase in house prices would be of only kr. 1.7 billion, even considering the larger OLS estimates.

While the yearly reduction in $CO_2$ emissions would be about five times larger, the economic value of these emission reductions provides a good reference point for the feasibility of such an investment.[18] In the best-case scenario, making these renovations incentive compatible in the short run would require a subsidy of about kr. 500 million. The achieved emission reductions would be currently priced at approximately kr. 39 million yearly, based on the EUA futures price (see footnote 18) and the EUR/DKK exchange rate.

It is beyond the scope of this paper to discuss whether such a subsidy would be efficient or optimal. Moreover, our simulations only consider houses that have been in the market: We cannot account for energy-inefficient houses strongly benefiting from energy renovations

[17] In this scenario we would be able to renovate about 25,000 houses, compared to the 36,000 houses renovated in the first scenario.

[18] We price each tonne of $CO_2$ according to the EU Emission Trading System futures price. At the time of performing this calculation (December 6 2021), the EUA futures price was at a record high of 81.25€.

which have not been in the market in the past decade. Nonetheless, our results highlight that emission reductions currently priced at approximately kr. 8 million yearly could be achievable even without explicit subsidies, as these are already incentive-compatible in the short run. This paper shows that not only identifying these renovations is possible, but also that such investments represent an easy first step in reaching established climate goals at no cost for the taxpayer.

### 5.2 Rising energy prices

Based on the discussion in section 4, we regard the price elasticities with respect to district heating prices (table 1) as being too uncertain. In our price scenario, we therefore focus on what would happen to house prices if the prices of other energy types increased. In light of the recent large increases in energy prices, such a scenario seems to become increasingly relevant to consider.

In a scenario for rising energy prices, there are fewer possible assumptions to make, relative to a scenario for renovation. As a point of departure, one may simply assume that the energy price increases by a certain percentage. As a baseline scenario, we assume a 20% increase in the kWh price.

Based on the fourth column in table 1 and our discussion in section 4, we conclude that increasing energy prices can be expected to lower house prices in the two least expensive areas, areas 0 and 1, respectively. A 20% increase in the energy price implies a reduction in house prices of 4.75% and 2.86% for price areas 0 and 1 respectively (correcting for the logarithmic approximation). Whether our results can be extrapolated to more extreme price increases is uncertain. For example, our analysis does not account for homeowners switching to other heating sources if their present one becomes sufficiently expensive.

## 6 Conclusions and implications for credit institutions and policy makers

Our analysis identifies both transitional risks and opportunities in the residential sector. Risks are primarily associated with rising energy prices and low return ratios for energy renovations and are concentrated in peripheral rural areas (e.g. Lolland, Southern and Western Jutland). In the past few years, such areas have been the center of political discussions in Denmark, with voiced concerns that financial institutions are overly prudent when it comes to financing real estate purchases in these areas. Though not directly related to those discussions, our analysis shows that these areas are among those that are most exposed to transitional climate risks.

Specifically, in these areas the already low housing values to be used as collateral by credit institutions are particularly exposed to rising energy prices, which recently have increased dramatically. For example, to the extent that the recent extraordinary large increases in the price of natural gas have increased expectations of future natural gas prices significantly, this is likely to have put a downward pressure on the value of homes in some of these areas.[19] Our scenario analysis further reveals that for an upcoming seller or a credit institute, there will typically be no incentive to finance these renovations as the investment costs are not going to be covered by the increase in the market value of the house. Financing

[19]For an analysis (in Danish) of the effect on sales prices of rising natural gas prices, see Ingholt and Møller (2022), which builds on the work underlying the present paper.

such investments can be risky, especially if the renovating households have already high loan-to-value ratios and little equity besides.

Nonetheless, two factors mitigate these risks. First, the investment amount from the energy report is an upper bound for the actual private cost for the homeowner, as these costs do not account for tax deductions and subsidies. For example, in Denmark it is possible to apply for a lump sum subsidy for energy efficiency improvements of one's home, depending on one's current energy rating, and until recently it has been possible to deduct expenses in connection with energy renovations. It is possible that similar tax deductions could be reintroduced. Moreover, the stated investment amount is based on the assumption that professional craftsmen are hired to carry out the improvements. Often proposals involving e.g. the replacement of insulation material or various minor components or devices can be carried out by homeowners themselves at a lower cost.

Second, not only may the investment cost be overestimated, our estimated price increases also reflect the current market valuation of energy efficiency. This valuation and the attention markets pay to energy efficiency might increase as energy prices increase, and energy consumption becomes more salient.

Due to these issues the return ratios for energy renovations computed in our scenarios could be regarded as lower bounds. Nevertheless, we still argue that the profitability of these renovations will still be lower, relatively speaking, in low-price rural and peripherical areas.

Financial opportunities consist in what we have referred to as a 'sweet spot' in terms of returns on energy renovation. For houses in the price range of kr. 1-2.5 million, typically located in and around smaller towns and areas with a higher population density, well-developed infrastructure, and road networks connected to towns and cities (e.g. Mid and Western Zealand, Funen and Eastern Jutland), returns on energy renovations are relatively high and often above 100% (see the map, panel b in figure 8). At the same time the investment cost associated with these houses is often relatively low. From the point of view of credit institutions, energy renovations are thus associated with little risk in these areas, as the collateral value can be expected to increase to at least match the issued loan for renovation.

When it comes to the opportunities for climate change mitigation, we have also shed some light on the potential for the residential sector to reduce $CO_2$ emissions. On the one hand, our results show that the renovations with the highest private returns are not those that will reduce $CO_2$ emissions the most. This finding suggests that private incentives can only partly contribute to the solution of the emission problem in this context. Specifically, there should be scope for policies that facilitate mitigation in the housing sector by, for example, means testing tax deductions and subsidies for renovations against the $CO_2$ reducing potential of the individual renovation. On the other hand, we have shown that it is possible to save over 13,000 tonnes of $CO_2$ emissions yearly, alone from renovations that are incentive-compatible in the sense that the investment costs are covered by the increase in sales price.

## References

ADOLFSEN, J., S. GUDIKSEN, A. T. MARTINELLO, B. M. MØNSTED, A. SCHMIDT, AND K. SONBERG (2022): "Segmentation of the housing market with internet data: Evidence from Denmark," Working Paper 188, Danmarks Nationalbank.

ANSELIN, L. (1988): *Spatial Econometrics: Methods and Models*, Studies in Operational Regional Science, Springer Netherlands.

ATHEY, S. AND G. IMBENS (2016): "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 113, 7353–7360.

ATHEY, S. AND G. W. IMBENS (2019): "Machine learning methods that economists should know about," *Annual Review of Economics*, 11, 685–725.

ATHEY, S., J. TIBSHIRANI, AND S. WAGER (2019): "Generalized random forests," *The Annals of Statistics*, 47, 1148–1178.

ATHEY, S. AND S. WAGER (2019): "Estimating treatment effects with causal forests: An application," *Observational Studies*, 5, 37–51.

BALLI, H. AND B. SORENSEN (2013): "Interaction effects in econometrics," *Empirical Economics*, 45, 583–603.

BATTEN, S., R. SOWERBUTTS, AND M. TANAKA (2016): "Let's talk about the weather: the impact of climate change on central banks," Bank of England working papers 603, Bank of England.

BATTOCCHI, K., E. DILLON, M. HEI, G. LEWIS, P. OKA, M. OPRESCU, AND V. SYRGKANIS (2021): "EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation," https://github.com/microsoft/EconML, version 12.0.

BINGLER, J. A. AND C. C. SENNI (2022): "Taming the Green Swan: a criteria-based analysis to improve the understanding of climate-related financial risk assessment tools," *Climate Policy*, 22, 356–370.

BROUNEN, D. AND N. KOK (2011): "On the economics of energy labels in the housing market," *Journal of Environmental Economics and Management*, 62, 166–179.

BRUEGGE, C., C. CARRIÓN-FLORES, AND J. C. POPE (2016): "Does the housing market value energy efficient homes? Evidence from the energy star program," *Regional Science and Urban Economics*, 57, 63–76.

CAN, A. AND I. MEGBOLUGBE (1997): "Spatial Dependence and House Price Index Construction," *The Journal of Real Estate Finance and Economics*, 14, 203–22.

CERIN, P., L. G. HASSEL, AND N. SEMENOVA (2014): "Energy performance and housing prices," *Sustainable Development*, 22, 404–419.

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, AND W. NEWEY (2017): "Double/debiased/neyman machine learning of treatment effects," *American Economic Review*, 107, 261–65.

CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.

DENG, Y., Z. LI, AND J. M. QUIGLEY (2012): “Economic returns to energy-efficient investments in the housing market: evidence from Singapore,” *Regional Science and Urban Economics*, 42, 506–515.

FERENTINOS, K., A. GIBBERD, AND B. GUIN (2021): “Climate policy and transition risk in the housing market,” Staff Working Paper 918, Bank of England.

FUERST, F., P. MCALLISTER, A. NANDA, AND P. WYATT (2015): “Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England,” *Energy Economics*, 48, 145–156.

FURUKAWA, K., H. ICHIUE, AND N. SHIRAKI (2020): “How Does Climate Change Interact with the Financial System? A Survey,” Bank of Japan Working Paper Series 20-E-8, Bank of Japan.

GELMAN, A. AND E. LOKEN (2014): “The statistical crisis in science: data-dependent analysis–a" garden of forking paths"–explains why many statistically significant comparisons don’t hold up,” *American scientist*, 102, 460–466.

GIBBONS, S. AND H. G. OVERMAN (2012): “MOSTLY POINTLESS SPATIAL ECONOMETRICS?*,” *Journal of Regional Science*, 52, 172–191.

HANSEN, A., O. JENSEN, AND J. KRAGH (2013): “Sammenhæng mellem energimærkning og salgspris.” *Netværk for energirenovering. SBi 2013: 06.*

HAYASHI, F. (2000): *Econometrics*, Princeton, N.J.: Princeton University Press.

HENDRY, D. AND B. NIELSEN (2007): *Econometric Modeling: A Likelihood Approach*, Princeton University Press.

HYLAND, M., R. C. LYONS, AND S. LYONS (2013): “The value of domestic building energy efficiency — evidence from Ireland,” *Energy Economics*, 40, 943–952.

INGHOLT, M. M. AND N. F. MØLLER (2022): “HØJERE GASPRISER KAN FØRE TIL LAVERE BOLIGPRISER I DELE AF LANDET,” Economic memo 4, Danmarks Nationalbank.

ISAKSON, H. (2002): “The Linear Algebra of the Sales Comparison Approach,” *Journal of Real Estate Research*, 24, 117–128.

JOHANSEN, S. (2006): “Confronting the economic model with the data,” in *Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, ed. by D. Colander, Cambridge University Press, vol. 1, chap. 15, 287–300.

KE, G., Q. MENG, T. FINLEY, T. WANG, W. CHEN, W. MA, Q. YE, AND T.-Y. LIU (2017): “Lightgbm: A highly efficient gradient boosting decision tree,” *Advances in neural information processing systems*, 30, 3146–3154.

KROGSTRUP, S. AND W. OMAN (2019): “Macroeconomic and Financial Policies for Climate Change Mitigation: A Review of the Literature,” IMF Working Papers 2019/185, International Monetary Fund.

KUMINOFF, N. V., C. F. PARMETER, AND J. C. POPE (2010): “Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?” *Journal of Environmental Economics and Management*, 60, 145–160.

LANG, G. AND B. LANZ (2022): “Climate policy without a price signal: Evidence on the implicit carbon price of energy efficiency in buildings,” *Journal of Environmental Economics and Management*, 111, 102560.

LAQUATRA, J., D. J. DACQUISTO, P. EMRATH, AND J. A. LAITNER (2002): "Housing market capitalization of energy efficiency revisited," *Summer Study on Energy Efficiency in Buildings. Washington, DC: American Council for an Energy-Efficient Economy*.

MIRONE, G. AND J. POESCHL (2021): "Flood risk discounts in the Danish housing market," Economic memo 7, Danmarks Nationalbank.

MURFIN, J. AND M. SPIEGEL (2020): "Is the risk of sea level rise capitalized in residential real estate?" *The Review of Financial Studies*, 33, 1217–1255.

NGUYEN, D. D., S. ONGENA, S. QI, AND V. SILA (2022): "Climate Change Risk and the Cost of Mortgage Credit*," *Review of Finance*, rfac013.

NÆSS-SCHMIDT, S., C. HEEBØLL, AND N. C. FREDSLUND (2015): "Do homes with better energy efficiency ratings have higher house prices?" *Report from Copenhagen Economics*.

PRIDE, D., J. LITTLE, AND M. MUELLER-STOFFELS (2018): "The value of residential energy efficiency in interior Alaska: a hedonic pricing analysis," *Energy policy*, 123, 450–460.

ROSEN, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

SCHUETZE, F. (2020): "Transition Risks and Opportunities in Residential Mortgages," DIW Berlin Discussion Paper 1910.

SMALL, K. A. AND S. S. STEIMETZ (2012): "SPATIAL HEDONICS AND THE WILLINGNESS TO PAY FOR RESIDENTIAL AMENITIES*," *Journal of Regional Science*, 52, 635–647.

SMITH, T. AND P. WU (2009): "A spatio-temporal model of housing prices based on individual sales transactions over time," *Journal of Geographical Systems*, 11, 333–355.

TARUTTIS, L. AND C. WEBER (2022): "Estimating the impact of energy efficiency on housing prices in Germany: Does regional disparity matter?" *Energy Economics*, 105, 105750.

THANOS, S., J. DUBÉ, AND D. LEGROS (2016): "Putting time into space: the temporal coherence of spatial applications in the housing market," *Regional Science and Urban Economics*, 58, 78–88.

VON GRAEVENITZ, K. AND T. E. PANDURO (2015): "An Alternative to the Standard Spatial Econometric Approaches in Hedonic House Price Models," *Land Economics*, 91, 386–409.

WAGER, S. AND S. ATHEY (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242.

WALLS, M., T. GERARDEN, K. PALMER, AND X. F. BAK (2017): "Is energy efficiency capitalized into home prices? Evidence from three US cities," *Journal of Environmental Economics and Management*, 82, 104–124.

# Appendix

## A Predeterminedness of $\overline{y}_{i,\tau}$

For consistency of OLS, we need regressors to be predetermined (i.e. not endogenous). This requires that they are uncorrelated with the current error term, here implying that $E[\overline{y}_{i,\tau}\varepsilon_{i,\tau}] = 0$, where $\varepsilon_{i,\tau}$ denotes the corresponding error term. To see what this restriction implies, assume first, without loss of generality, that the average, $\overline{y}_{i,\tau}$ simply includes only one past sales price and ignore regressors other than a constant. Hence, the regression equation becomes, $y_{i,\tau} = \alpha + \beta\overline{y}_{i,\tau} + \varepsilon_{i,\tau}$, where $y_{i,\tau}$ is the log sales price of property $i$ sold at time $\tau$ and $\overline{y}_{i,\tau} = y_{j,\widetilde{\tau}}$, i.e. the sales price of property $j$ sold at time $\widetilde{\tau} < \tau$. It follows that $E[\overline{y}_{i,\tau}\varepsilon_{i,\tau}] = E[y_{j,\widetilde{\tau}}\varepsilon_{i,\tau}] = E[(\alpha + \beta\overline{y}_{j,\widetilde{\tau}} + \varepsilon_{j,\widetilde{\tau}})\varepsilon_{i,\tau}] = \beta\text{Cov}(\overline{y}_{j,\widetilde{\tau}}; \varepsilon_{i,\tau}) + \text{Cov}(\varepsilon_{j,\widetilde{\tau}}; \varepsilon_{i,\tau})$. Therefore, since $\overline{y}_{j,\widetilde{\tau}}$ depends on previous error terms, the predeterminedness of the regressor $\overline{y}_{i,\tau}$ requires zero correlation between the current and past error terms. This result stresses that in constructing $\overline{y}_{i,\tau}$, care is taken to ensure that this measure includes enough 'comparable sales' by specifying a sufficient radius and going enough back in time, otherwise the error terms will correlate so that $E[\overline{y}_{i,\tau}\varepsilon_{i,\tau}] \neq 0$.

## B Data description

The data we consider consist of daily observations of realized sales prices and determinants of these collected for Denmark over the period January 1 2014 to November 18, 2020. Each observation corresponds to a property sold only once during this period. The sources are described in section 3. Table 4 contains a grouped list of all variables included in the analysis and their main summary statistics.

#### Regressand

The regressand is the natural logarithm of the sales price of property $i$.

#### Main regressors

**Standardized energy consumption:** The natural logarithm of standardized energy consumption in kWh per square meter of heated area per year. Hence, this is the reciprocal of energy efficiency. The current Danish energy rating scheme, with labels from A2020 to G, is based directly on intervals for the standardized energy consumption per square meter per year. This consumption is however only associated with heating space and water. As mentioned in the text above, we include total electricity consumption. For further details, see above.

**Energy price:** The natural logarithm of a weighted average of energy prices across all energy sources used in the property. The weights are the respective shares of standardized consumption in kWh. For further details, see above.

#### Control variables

In the attempt to recover a causal effect from the main regressors on sales prices, it is important to control for a wide range of available variables, particularly those that are expected to be important for the sales price and likely also correlate with efficiency and/or energy prices (i.e. confounders). One obvious aspect to control for is the physical quality or condition of the house. In particular, properties with low energy efficiency are typically also properties in a poorer general condition. Quality is a many-faceted and intangible variable, and we proxy this by the data from the official condition reports for the property (damages, defects, and omissions). We have access to reports both for the general condition and electrical installations of the property. In particular, from both reports we included the number of so-called K1s, K2s and K3s, which cover damage, defects, or omissions that can be classified as less severe, critical and critical with potential spreading to other building parts, respectively. In addition, we also included the number of 'issues' that the expert, during his/her inspection, suggested should be examined further. These are the 'unknown concerns' in table 4.

We included square meter measures for the inhabital area, total built-up area and plot area.

A range of other technical/structural characteristics of the house were also included: In particular, we included indicator variables for the type of roof, the number of rooms, the number of bathrooms, the number of floors, the type of main heating source, the type of supplementary heating source, whether the property includes a garage, a shed, and a covered terrace.

**Table 4**
**Descriptive table, continuous variables**

| Variable name | Mean | Median | Std. deviation | Pct. missing | Histogram |
|---|---|---|---|---|---|
| | | | *Regressand* | | |
| Log price per $m^2$ | 9.5 | 9.5 | 0.7 | 0.0 | |
| | | | *Main regressors* | | |
| Log energy consumption per $m^2$ | 5.2 | 5.2 | 0.4 | 0.0 | |
| Log energy price | -0.1 | -0.1 | 0.3 | 0.0 | |
| | | | *Continuous regressors* | | |
| Inhabitable $m^2$ | 141.5 | 136.0 | 41.5 | 0.0 | |
| Neighborhood price per $m^2$ ($\overline{p}_{i,\tau}$) | 15808.2 | 13188.0 | 9514.3 | 1.5 | |
| Plot size $m^2$ | 772.6 | 788.0 | 317.3 | 0.0 | |
| Meters above sea level | 27.8 | 23.3 | 20.8 | 0.1 | |
| Distance to coastline (meters) | 7331.4 | 4384.8 | 8631.4 | 0.2 | |
| | | | *Ordinal regressors* | | |
| Number of rooms | 4.8 | 5.0 | 1.3 | 0.0 | |
| Number of bathrooms | 1.4 | 1.0 | 0.5 | 0.0 | |
| Construction year | 1961.7 | 1967.0 | 29.8 | 0.0 | |
| Number of floors | 1.1 | 1.0 | 0.2 | 0.0 | |
| Number of buildings | 2.2 | 2.0 | 0.8 | 0.0 | |
| Building age at sale | 56.2 | 50.0 | 33.2 | 0.0 | |
| Minor damage | 12.2 | 11.0 | 8.6 | 11.1 | |
| Major damage | 7.3 | 6.0 | 7.0 | 11.1 | |
| Major damage with risk of spreading | 4.7 | 4.0 | 4.7 | 11.1 | |
| Unknown concerns in status report | 0.1 | 0.0 | 0.4 | 11.1 | |
| Minor damage (electricity) | 6.5 | 5.0 | 5.1 | 18.2 | |
| Major damage (electricity) | 2.5 | 2.0 | 3.1 | 18.2 | |
| Major damage with risk of spreading (electricity) | 0.5 | 0.0 | 0.8 | 18.2 | |
| Unknown concerns in electricity report | 0.6 | 0.0 | 1.7 | 18.2 | |

**Table 5**
**Descriptive table, categorical variables**

| Variable name | Mean | Median | Std. deviation | Pct. missing | Histogram |
|---|---|---|---|---|---|
| | | | *Categorical regressors* | | |
| Type of house (detached, terraced) | | | | 0.0 | |
| Heating source | | | | 0.0 | |
| Supplementary heating source | | | | 0.0 | |
| Roof type | | | | 0.0 | |
| Garage dummy | | | | 0.0 | |
| Auxiliary shed dummy | | | | 0.0 | |
| Neighborhood size for $\overline{p}_{i,\tau}$ | | | | 1.5 | |
| | | | *Fixed effects* | | |
| Municipality | | | | 0.0 | |

## C Estimation assuming homogenous sales price effects of energy efficiency and prices

Table 6 shows the estimates of the house price elasticities with respect to the two main regressors, standardized energy consumption and energy price, based on a model assuming homogenous effects. Hence, these estimates do not account for heterogeneity. They are meant to serve as benchmark estimates of the average effects to which we can compare the estimates which take heterogeneity into account (table 1).

We consider four different estimations: The first row in table 6 contains the results of a linear regression of sales price on standardized energy consumption and energy price, all in logs, estimated with OLS with no controls (other than a constant). In the second row this is augmented to include all controls except the geographical ones (price area, municipalities and average neighborhood price level). In the third row the geographical controls are further added so this is the preferred estimation among the first three. Finally, in the fourth row all controls are included (as in the third row), and the estimation is based on the double machine learning approach described in section 2.2. Inference is provided in the form of standard t-ratios. For the preferred estimation and the estimation with all controls except the geographical ones, we also report t-ratios based on clustering the variance-co-variance structure by municipality.

The results in table 6 suggest that lower energy efficiency (higher standardized consumption) reduces house prices. There are substantial differences in the order of magnitude of the estimates across the four estimations. Comparing the simple unconditional estimation in the first row with the preferred ones in the third and fourth rows stresses the need of adjusting for structural and geographical variables. The first stages of the DML models further support this interpretation. For these models we obtain an out-of-sample $R^2$ of 84% for house prices, 71% for standardized energy consumption, and 74% for energy prices. In other words, our controls account for the great majority of the variation in each of our variables of interest.

The preferred estimate based on OLS is -0.236% and thus twice as big as that based on DML (-0.12%). The two estimates represent an upper and a lower bound for the average (absolute) elasticity. The difference arises as the non-linear orthogonalization in the DML estimation controls for confounding variation to a larger extent.

For the elasticity with respect to energy price, the results are primarily negative, but smaller and less significant. However, focusing on the preferred specifications in rows 3 and 4, the estimates are negative and significant (borderline significant when clustering). The counter-intuitive positive sign in the second row is most likely due to unobserved heterogeneity and is discussed in section 4.

**Table 6**
**Estimated aggregate price elasticities (homogenous effects)**

| | Standardized energy consumption | | | Energy price | | |
|---|---|---|---|---|---|---|
| | Point estimate | t-value | t-value (clust.) | Point estimate | t-value | t-value (clust.) |
| OLS, no controls | -0.439 | -95.262 | | -0.010 | -1.397 | |
| OLS, no geographic controls | -0.054 | -10.841 | -0.942 | 0.170 | 21.334 | 1.595 |
| OLS | -0.236 | -85.598 | -13.367 | -0.038 | -8.227 | -1.926 |
| DML | -0.120 | -32.643 | | -0.056 | -10.439 | |

NOTES: All results are based on a common sample of 195,395 non-repeated individual sales of single family homes and townhouses in Denmark taking place between January 2014 and May 2020. The table shows coefficients and t-values for both standardized energy consumption and energy price, where the regressand is the log square meter prices at sale. The first and fourth columns show point estimates; the second and fifth columns show t-values computed under the assumption of independent observations; The third and sixth columns show t-values computed by allowing for arbitrary autocorrelation of errors within municipalities. The rows indicate the model and the controls used for the estimation. For a full list of controls included in the regression see tables 5 and 4.

## D Causal forests: Additional results and diagnostics

**Table 7**
**Omnibus test for heterogeneity**

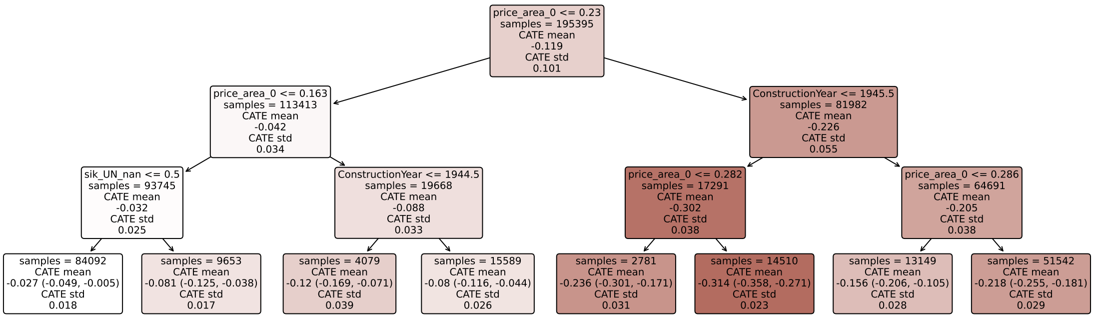
| | Point estimate | Std. error | t-value | 95% confidence interval | |
|---|---|---|---|---|---|
| | | | *Standardized consumption* | | |
| Mean prediction ($\beta_C$) | 0.896 | 0.029 | -3.546 | 0.839 | 0.954 |
| Differential prediction ($\beta_D$) | 1.335 | 0.040 | 13.575 | 1.678 | 1.907 |
| | | | *Energy price* | | |
| Mean prediction ($\beta_C$) | 0.811 | 0.137 | -1.377 | 0.543 | 1.080 |
| Differential prediction ($\beta_D$) | 1.183 | 0.056 | 2.274 | 1.086 | 2.160 |

NOTES: The table shows the results of the omnibus test for heterogeneity suggested by Athey and Wager (2019), which checks whether the estimated Conditional Average Treatment Effects (CATEs) predict well (out of sample) the orthogonalized outcome, given the orthogonalized regressor. Denote $\hat{\tau}^{-i}$ as the estimated CATE (the $^{-i}$ superscript means that the CATE is estimated out-of-bag), $\bar{\tau}$ as the average estimated CATE, $y_i - \hat{m}^{-i}(X_i)$ the residualized outcome, where $\hat{m}^{-i}$ is the estimated first-stage (set of) model(s), and $T_i - \hat{e}^{-i}(X_i)$ the residualized regressor.
We then estimate the regression

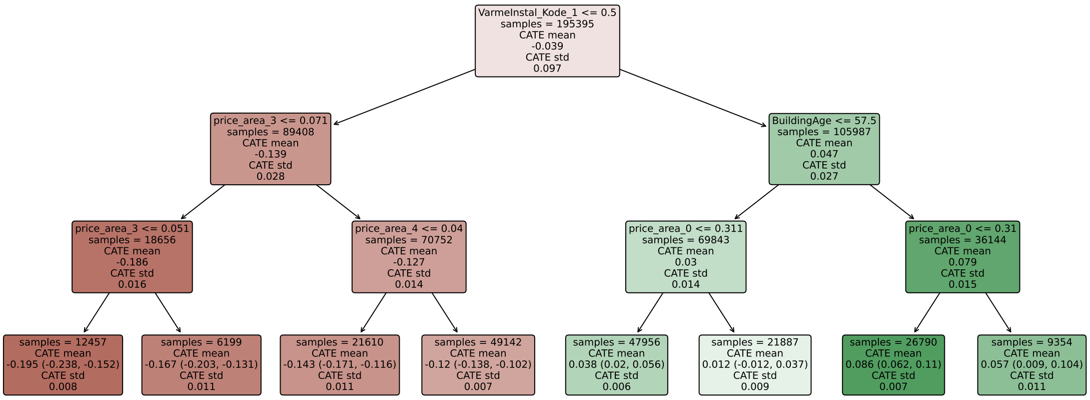
$$y_i - \hat{m}^{-i}(X_i) = \beta_C \cdot \underbrace{\bar{\tau}\left(T_i - \hat{e}^{-i}(X_i)\right)}_{C_i} + \beta_D \cdot \underbrace{\left(\hat{\tau}^{-i} - \bar{\tau}\right)\left(T_i - \hat{e}^{-i}(X_i)\right)}_{D_i}$$

where the coefficient $\beta_C$ represents how well the CATEs capture the average effect, and $\beta_D$ captures how well the causal forest capture heterogeneity. In a well calibrated model, both coefficient should be equal to 1. As Athey and Wager argue, the p-value of $\beta_D$ can act as a test of the hypothesis that the causal forests found meaningful heterogeneity in the data.
This test reveals that while we detect substantial heterogeneity in the data, the Causal Forest model is not perfectly calibrated, and therefore we use it only as a guidance for identifying relevant dimensions of heterogeneity.

**Figure 10**
**Interpreting CFs**



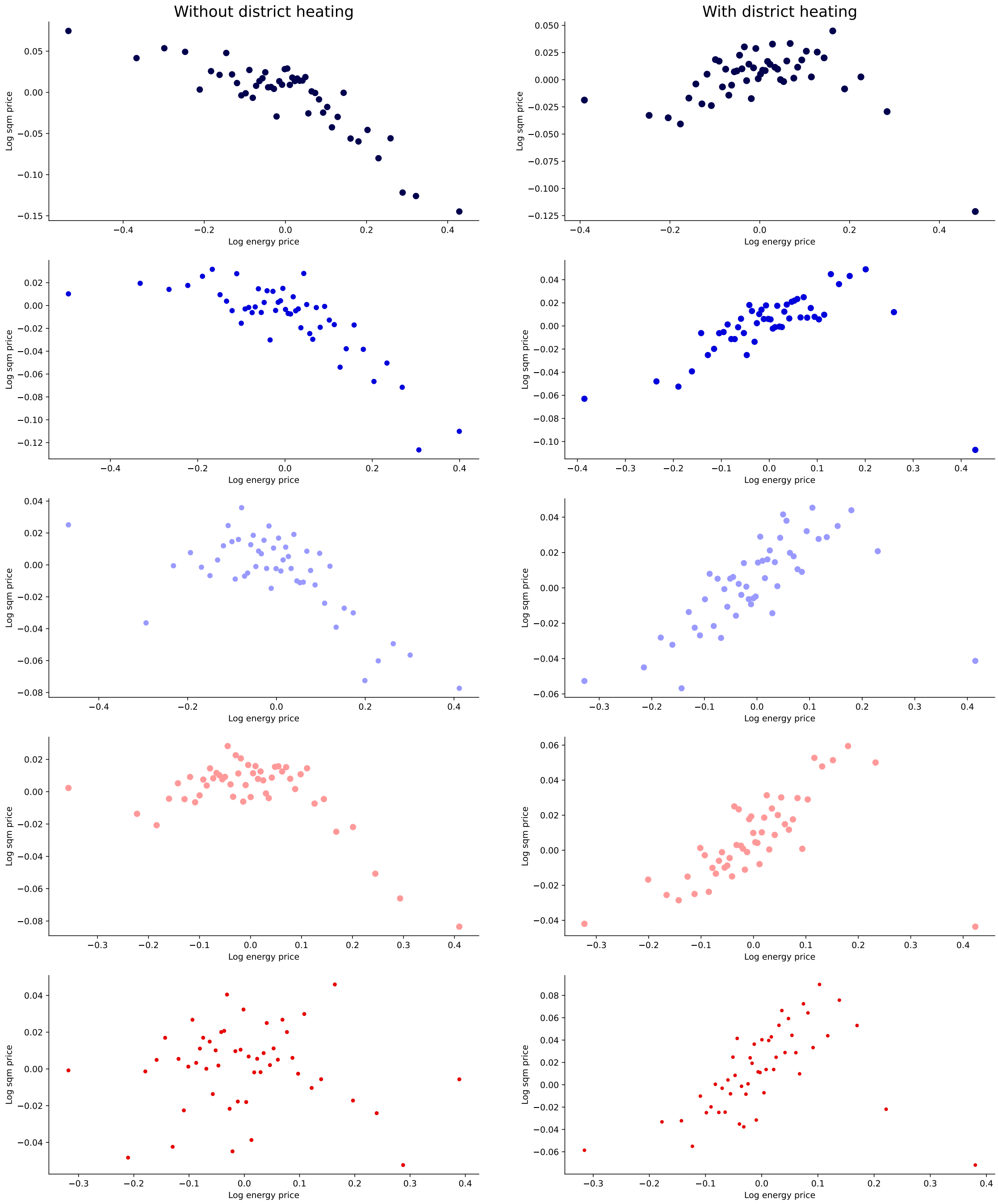

**(a) Energy consumption**




**(b) Energy price**

# E Additional results

**Figure 11**
**Partial regression plots for energy prices by price area and type of heating**




NOTES: The figure shows a binned residualized scatterplot of the relationship between log house prices per m2 and key variables of interest, divided by price areas (left) and heating source (right). Each plot residualizes the main regressor of interest and the regressand using the estimated first-stage gradient boosted trees models. For each price area we then bin the residuals in 50 approximately equally-sized bins according to the regressor residuals, such that each bin represents approximately 2% of the data within each group. Finally, we plot the average of the regressand and regressor residuals against each other for each bin in a scatterplot, where the size of each point is proportional to the number of observations it represents.

PUBLICATIONS

## NEWS

News offers quick and accessible insight into an Analysis, an Economic Memo, a Working Paper or a Report from Danmarks Nationalbank. News is published continuously.

## ANALYSIS

Analyses from Danmarks Nationalbank focus on economic and financial matters. Some Analyses are published at regular intervals, e.g. *Outlook for the Danish economy* and *Financial stability*. Other Analyses are published continuously.

## REPORT

Reports comprise recurring reports and reviews of the functioning of Danmarks Nationalbank and include, for instance, the *Annual report* and the annual publication *Danish government borrowing and debt.*

## ECONOMIC MEMO

An Economic Memo is a cross between an Analysis and a Working Paper and often shows the ongoing study of the authors. The publication series is primarily aimed at professionals. Economic Memos are published continuously.

## WORKING PAPER

Working Papers present research projects by economists in Danmarks Nationalbank and their associates. The series is primarily targeted at professionals and people with an interest in academia. Working Papers are published continuously.

DANMARKS NATIONALBANK
LANGELINIE ALLÉ 47
DK-2100 COPENHAGEN Ø
WWW.NATIONALBANKEN.DK

As a rule, Working Papers are not translated, but are available in the original language used by the contributor.

Danmarks Nationalbank's Working Paper are published in PDF format at www.nationalbanken.dk. A free electronic subscription is also available at the website. The subscriber receives an e-mail notification whenever a new Working Paper is published.



Please direct any enquiries directly to the contributors or to Danmarks Nationalbank, Communications, *Kommunikation@nationalbanken.dk.*

DANMARKS NATIONALBANK