BANK OF ENGLAND

Staff Working Paper No. 918
Climate policy and transition risk in the housing market

Konstantinos Ferentinos, Alex Gibberd and Benjamin Guin

April 2021

BANK OF ENGLAND

Staff Working Paper No. 918

Climate policy and transition risk in the housing market

Konstantinos Ferentinos,[1] Alex Gibberd[2] and Benjamin Guin[3]

Abstract

Public policies aimed at mitigating climate change can come with the transition risk of sudden adjustments of asset prices. We study the consequences of a policy intervention addressing greenhouse gas emissions in the housing market. Leveraging a unique data set of the population of all house transactions in England and Wales, we document novel evidence of transition risk.

Prices of carbon-intensive properties affected by this policy decreased by about £5,000 to £9,000 relative to unaffected ones. We interpret this result as evidence in favour of semi-strong market efficiency in the housing market. We infer moderate implications for financial stability and for the wealth distribution among homeowners.

Key words: Climate policy, transition risk, house prices, financial stability, wealth inequality.

JEL classification: C54, Q54, Q58.

---

(1) Lancaster University. Email: ferentinos.konstantinos.msc@gmail.com
(2) Lancaster University. Email: a.gibberd@lancaster.ac.uk
(3) Bank of England. Email: benjamin.guin@bankofengland.co.uk

We are grateful for comments from Thiemo Fetzer, Bill Francis, Simon Hall, Daire McCoy, Jennifer Owladi, Martin Sheppard and a reviewer of the Staff Working Paper Series of the Bank of England. Any views expressed are solely those of the authors and should not be taken to represent those of the Bank of England or as a statement of Bank of England policy. This paper should therefore not be reported as representing the views of members of the Monetary Policy Committee, Financial Policy Committee or the Prudential Regulation Committee.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Residential housing accounts for a significant share of greenhouse gas emissions, about 15% in the United Kingdom[1], which is a significant contribution to climate change. Public policies have been proposed to mitigate this effect but, they in turn, can potentially adversely affect property prices. Such adjustments of asset prices are often referred to as "transition risk" from climate change (Carney, 2015) which can undermine financial stability. Moreover, price declines can erode voters' willingness to support such policies and governments' inclination to implement them. To the best of our knowledge, evidence on the adjustment of house prices to climate policies is non-existing. This lack of evidence is partly due to scarce data and the fact that transition risk might materialize only in the future once climate policies will be implemented.

This paper addresses this gap in the context of a specific policy intervention in the UK housing market which targeted houses that were the least energy-efficient. This Minimum Energy Efficiency Standard (MEES) which came into force in England and Wales on 1 April 2018 aimed at encouraging landlords and property owners to improve the energy efficiency of their properties which should reduce overall greenhouse gas emission. It restricted the granting and continuation of existing tenancies where the property was not energy-efficient, with an Energy Performance Certificate (EPC) rating of F and G. Otherwise landlords would face a fine of up to £5,000.

In this paper, we examine how sales prices of properties adjust to the implementation of this policy. We expect prices of the least energy-efficient properties which are affected by the policy to decrease relative to properties that are unaffected. To test this hypothesis, we construct a novel data set of the population of all property transactions in England and Wales which we match with the Energy Performance Certificate (EPC) ratings of these buildings. The panel dimension of our data allows us to track the prices of the same properties over time. This mitigates concerns around confounding factors such as economic conditions or the composition of the housing stock changing over time.

We conduct a Difference-in-Difference (DiD) analysis which we combine with propensity score matching (PSM). As a first step, we conduct the PSM to ensure that we compare only properties that are like-for-like. We match each property affected by the MEES 2018 with a "twin" property, a different property that is otherwise very similar based on observable characteristics, such as the property size or the region it is located in, but unaffected by the policy. We then gauge the effect of the policy on property prices, by conducting a Difference-in-Difference (DiD) analysis in which we compare how transaction prices of properties affected by this policy change relative to properties that are unaffected by the policy.

Our results suggest that prices of properties affected by the MEES 2018 decreased by about £5,000 to £9,000 relative to unaffected ones. The magnitude of this effect compares well to our priors. As energy performance certificates are publicly available, we expect potential buyers to consider them when making their purchase decisions. The degree of the price decrease should reflect the expected cost necessary to improve the energy efficiency of the property[2] and the possible fine if a landlord does not comply. Costs are

[1] Source: https://www.gov.uk/government/statistics/final-uk-greenhouse-gas-emissions-national-statistics-1990-to-2018, accessed on 09 November 2020

[2] EPC certificates also contain recommended measures and their anticipated costs. Link to a sample EPC cer-

initially capped at £3,500 but homeowners will need to incur the remaining costs and improve the energy efficiency of the property after 5 years. If landlords, i.e. the owners of properties that let them out, are unwilling to improve the energy efficiency, they will face a fine up to £5k and will need to incur the cost of improvement anyway. Hence, we argue a price discount in the range of £5k to £9k might be in an order of magnitude that we could expect. Overall, we take this result as evidence in favour of semi-strong market efficiency (Fama, 1970), as real estate markets seem to price in the publicly available information about the energy efficiency of the underlying property.

We also discuss to what extent this policy had implications for both financial stability and the wealth distribution of homeowners in England and Wales. First, decreases in property prices mean that collateral values of outstanding mortgages decrease. Once outstanding loan amounts exceed the value of collateral, mortgage lenders would incur losses if borrowers default on their mortgage payments. To gauge the severity of this concern, we conduct a back-of-the envelope calculation. Using a proprietary data set on the stock of mortgages outstanding in year-end 2017, we calculate the share of mortgages whose outstanding loan amount is at least 90% of the property value. We then subtract our best estimate of the price discount of about £9k from the least energy efficient properties with Energy Performance Certificate ratings of F and G. This back-of-the envelope calculation suggests that the MEES 2018 pushed an additional 0.5% of mortgages outstanding in end 2017 into the range of high outstanding loan-to-value ratios of at least 90%.[3] As a result, the share of such high LTV mortgages increases by 2.8%. We take this as evidence that implications for financial stability are limited.

Second, climate policies might have implications for the wealth distribution among homeowners in the UK. Before the MEES 2018 policy intervention, values of energy-efficient properties were on average higher than values of inefficient properties. The MEES 2018 should have decreased the property values, and, hence, the wealth of the owners of the least energy-efficient properties relative to owners of more energy-efficient ones. Everything else equal, this MEES 2018 should have increased wealth inequality among homeowners. To quantify this change, we calculate the Gini coefficient, an established measure of statistical dispersion intended to represent the wealth inequality within a country (Fagereng et al., 2016) where lower values indicate higher equality. We then compare the Gini coefficient among homeowners before the MEES 2018 with the Gini coefficient after this policy intervention. We observe that the Gini coefficient in our sample slightly decreases by about 1 percentage point since the implementation of the MEES 2018, i.e. wealth equality among UK homeowners increased. To gauge the effect of the MEES 2018, we then construct a counterfactual, the Gini coefficient if the MEES 2018 had not been implemented. This counterfactual analysis suggests that, in absence of the MEES 2018, the Gini coefficient would have decreased by another 1 percentage point further. We take this as evidence that the MEES 2018 policy intervention might have led to moderate increases in the wealth inequality among UK homeowners.[4]

tificate: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/5996/2116821.pdf`, accessed on 09 November 2020.

[3]This may be a range if we assume that a bank incurs costs to repossess the property and sell the property, possible below market value.

[4]This conclusion holds under strong assumptions. For simplicity, we assume that each homeowner owns one property as we cannot identify homeowners with multiple homes in the data.

The present paper contributes to three strands of the literature. First, we document how a climate policy can lead to adjustments of asset prices, in this case transaction prices of houses. While there is some evidence that transition risk from climate change is capitalized in the stock market (Hong et al., 2019; Krueger et al., 2020; Bolton and Kacperczyk, 2020; Sautner et al., 2020), evidence on the real estate market is scarce. The present paper addresses this gap. Examining the MEES 2018 policy, we find a negative effect on transaction prices in an order of magnitude to be expected from the calibration of fines and exemptions of the policy. By providing such evidence, the present paper also contributes to a fundamental question in economics and finance whether and to what extent markets are efficient. Going back to Mandelbrot (1963) and Samuelson (1965), there has been an ongoing and unresolved debate whether publicly available information are reflected in asset prices, something Fama (1970) called semi-strong information efficiency. Empirical evidence on the efficiency of the real estate market is ambiguous but it casts doubt on its efficiency (Case and Shiller, 1989, 1990).[5] The MEES 2018 offers an ideal quasi-natural experimental set-up to revisit the question of efficiency of real estate markets: First, energy performance certificates are publicly available and, hence, accessible to all market participants. Second, by setting fines and the value of exemptions from the policy, there are clear lower bounds on the expected effect on transaction prices. Hence, we have a clear theoretically-founded empirical prior of the order of magnitude of the effect of the policy on property prices which is testable.

For our empirical analysis, we employ a Difference-in-Difference design. Its validity relies on one key identifying assumption, the parallel trends assumption. It requires that in the absence of the policy intervention, the difference in the average of the outcome variable, in our case the transaction price of the property, between properties affected by the policy and the ones unaffected by the policy would have evolved in parallel (e.g. Lechner, 2010). Researchers often assess the plausibility of this parallel trends assumption by examining the pre-trends, i.e. the differences in trends between the treatment and control group prior to the policy intervention. Standard approaches to verifying this assumption do not typically go beyond visual inspections or testing for differences in linear pre-intervention trends (Ryan et al., 2019; Roth, 2020; Rambachan and Roth, 2019). In fact, out of the 16 papers in the 2016 American Economic Review that use a linear panel data model, eleven are concerned with the existence of pre-trends as a sign of endogeneity. Of these papers, 9 include a plot of pre-trends, of which provide a formal test of whether pre-trends are zero (Freyaldenhoven et al., 2019). Given the importance of this assumption for the validity of the DiD inference, any visual conclusion should ideally be verified under a statistical framework. Moreover, statistical tests for differences in pre-intervention outcome trends between treated and non-treated units, should not make overly restrictive assumptions about the functional form of the trend. Instead, a more realistic approach is to avoid the assumption of a linear relationship between outcome and time, and allow the observed data to determine the shape of the trend in outcomes. To this end, we here present a novel application of *Generalized Additive Models (GAMs)* (Hastie and Tibshirani, 1986), allowing for smooth non-linear trends over time.

The remainder of the paper is as follows: Section 2 summarises the policy background; section 3 offers a conceptual framework. Section 4 details the data used in our analysis; section 5 presents the key methodology used, alongside results. Last, section 6 provides an economic interpretation of the results and conclusion.

[5]See Herath and Maier (2015) for a meta-analysis of the literature.

2 Policy Background

The 2018 Minimum Energy Efficiency Standard initiative (MEES 2018) came into force in England and Wales on 1 April 2018 under the Energy Efficiency (Private Rented Property) Regulations 2015 (UK Statutory Instruments, 2015). It is best understood as a part of a large-scale market transformation strategy, that uses the EPCs as a springboard to transform the overall UK housing market and reduce greenhouse gases emissions due to the building stock's heavy energy use.

The policy supported two statutory objectives of UK government. First, it is aimed at reducing energy demand and greenhouse gas emissions. By improving the energy efficiency of privately rented homes, this policy would cut energy use and the greenhouse gas emissions, contributing to the government's climate change commitments. Second, it also was intended to tackle fuel poverty. Raising energy efficiency standards to an EPC rating of E by 2020, mirrors the government's interim target to raise as many fuel poor homes in England to energy efficiency Band E by the same date (Department for Business, Energy & Industrial Strategy, 2018).

The policy aimed to address a range of market failures and barriers to energy efficiency improvements which provided a rationale for government intervention in the private rental market. First, there was a concern of misaligned incentives as the costs of upgrading a property fall to landlords but the benefits of lower energy costs and/or a warmer home accrue to the tenant. The landlord was not necessarily able to capture the benefits through increases in rent. Second, there were possible externalities, such as energy prices not fully reflecting the climate change costs of burning fossil fuels, or the public health benefits of warmer homes not fully accruing to those who pay for energy efficiency upgrades. Third, there was a concern of incomplete information as landlords or tenants might not have a good understanding of the benefits of energy efficiency (Department for Business, Energy & Industrial Strategy, 2018).

The MEES 2018 initiative means that, since 1 April 2018, landlords of domestic private rented properties in England and Wales are prohibited to let a tenancy to new or existing tenants if the property that is to be let has an EPC rating of F or G (Department for Business, Energy & Industrial Strategy, 2017, p. 11). This minimum standard takes effect from the point at which a new tenancy is issued, or where an existing tenancy is renewed.[6] For properties that have an EPC rating below E, the landlord needs to improve the energy performance of the property in order to ensure that the standard of EPC E is met or even better exceeded (Department for Business, Energy & Industrial Strategy, 2017, p. 10). An exemption is allowed, in situations where the landlord cannot increase the rating to E or above at costs lower than £3,500 (Department for Business, Energy & Industrial Strategy, 2017, p. 57). The exemption remains valid for five years. After that, the landlord must implement the required improvements of the property, so as to comply with the policy. In cases where the property does not meet the criteria for an exemption and the landlord does not comply with the policy, they may face a penalty of up to £5,000 (Department for Business, Energy & Industrial Strategy, 2017, p. 89).

[6]The standard will later be expanded to all relevant properties, even where there has been no change in tenancy, from 1 April 2020 in the domestic sector.

3 Conceptual Framework

In this section, we provide a conceptual framework which discusses the implications of this policy for the housing market. We take the view of a prospective property buyer of an energy-inefficient property and ask how their willingness to pay decreases in light of the policy intervention given the magnitude of exemptions and penalties. We then show the implications for transaction prices.

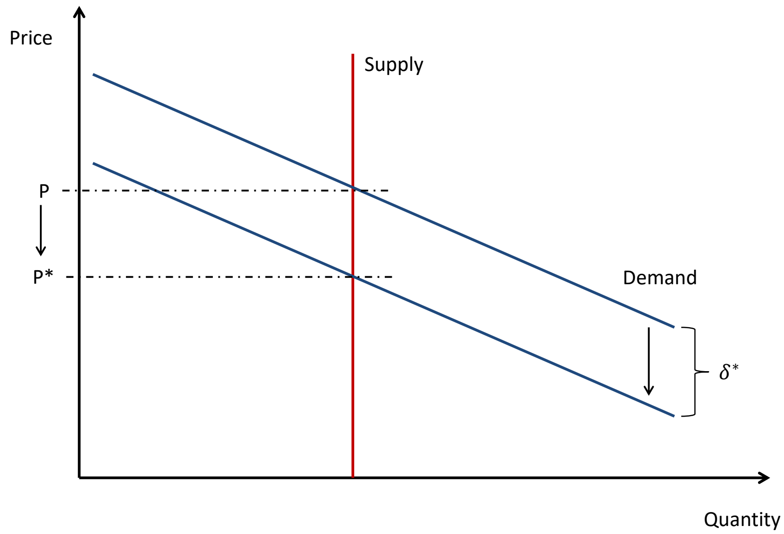
First, we discuss how the MEES 2018 should have affected a buyer's willingness to pay for a property. For that, consider a risk-neutral buyer who wants to ultimately let the property. This buyer considers purchasing a property with an EPC rating of below E. Initially, the costs of home improvements are capped at £3,500 by the policy. However, the buyer must still conduct the remaining home improvement after 5 years. Hence, their overall willingness to pay for the property should decrease by £3,500 plus the remaining cost of improvements in five years discounted at a rate of $\rho \geq 0$. Alternatively, this buyer can choose not to invest in the property to increase its energy efficiency. In this case, they might be caught with a probability of $0 \leq \pi \leq 1$ and will need to pay the fine of up to £5,000.[7] The buyer will need to incur the cost for the home improvement anyways. Hence, their willingness to pay should decrease by the probability of getting fined times the fine plus the costs of improvement. The willingness to pay δ^* of a risk-neutral buyer will decrease by the minimum of the expected costs of improvement or the expected costs of not improving with the risk of getting fined.

$$\delta^* = \min \begin{cases} 3500 \ + \ (\frac{1}{1+\rho})^5 (\ Cost \ - \ 3500 \ ) \\ \pi(\ 5000 \ + \ Cost \ ) \end{cases} \tag{1}$$

How does this affect this change in a buyers willingness to pay affect transaction prices? For simplicity, we assume the supply properties to be perfectly inelastic, i.e. it does not change with price. This may be a reasonable assumption for a relatively short time period which we study as houses or flats cannot be easily built. In this case, the decrease in a buyer's willingness to pay should be directly reflected in transaction prices, i.e. the decrease in the equilibrium price, or transaction price from P to P* should be equal to the change of a buyer's willingness to pay δ^* (see in Figure 1).

[7]For simplicity, we assume that the full fine of £5k applies.

Figure 1: Effect of the MEES 2018 on transaction prices of energy-inefficient properties



This simple conceptual framework is clearly not exhaustive. In particular, it discusses the implications for the private rented housing market. It may not hold for the owner-occupied market where buyers may not be interested in letting out the property. However, even in the absence of direct penalty, the introduction of interventions such as MEES 2018 may motivate property purchasers to invest in improvements as they may set a precedent for further energy efficiency interventions. Thus, even if the intervention does not directly impact them (via equation 1) they may be encouraged to invest in improvements as a spillover effect.

4 Data

4.1 Data sources

In order to conduct an empirical assessment of the MEES 2018 policy, we combine data from three different sources. These data cover information on transaction of properties, the energy performance of these properties as well as demographic characteristics of the regions in which they are located. We summarise these data sets below. We provide additional details in Appendix A.

Data on property transactions First, we derive a data set which includes information on property transactions, sourced from HM Land Registry (HM Land Registry, 2014). This data set includes information on all residential property transactions in England and Wales since 1995. In particular, it includes information on the date of the transaction and the price paid as well as the exact address of each property. Our empirical strategy requires data on transactions from both before and after the MEES 2018 policy intervention. For that reason, we focus only on those properties which have multiple transactions, both before and after the intervention date.[8] Specifically, our final sample consists of such properties which were repeatedly transacted in the time period between 2015 and 2019.

Data on Energy Performance Certificates (EPC) To identify the properties which are affected by the MEES 2018 policy intervention, we use data from the public register on Energy Performance Certificates (EPC) which is the official source for all EPCs issued for all domestic buildings and building units in England and Wales. This register covers information on all properties that have been constructed, sold or let since 2008, and were sourced from the Ministry of Housing, Communities & Local Government (MHCLG) (Ministry of Housing, Communities & Local Government, 2020a). EPCs are a good proxy of energy efficiency but can come with some measurement error (Hardy and Glew, 2019; Crawley et al., 2019).[9]

Data on geodemographic characteristics We complement these data with information characterising the area surrounding each property, making use of geodemographic classifications in England and Wales. Specifically, we use the classification produced from the 2011 census which clusters communities into eight different types (Gale et al., 2016): 1) Rural Residents; 2) Cosmopolitans; 3) Ethnicity Central; 4) Multicultural Metropolitans; 5) Urbanites; 6) Suburbanites; 7) Constrained City Dwellers; 8) Hard-Pressed Living.[10] These data are at the output area (OA) level, with each OA consisting of one type of postcode units, either entirely urban or entirely rural postcodes. OAs are designed to share similar population sizes, and attain a high level of social homogeneity, with respect to the type of tenure and accommodation, with a recommended size of 125 households.[11] We sourced these data from the Open Geography Portal (Office for National Statistics, 2011).

While the MEES 2018 policy intervention applies specifically to domestic private rented properties, the analysis in this paper examines all repeated property sales in England and Wales, that are sold for value and are lodged with HM Land Registry for registration. The rationale behind this decision is that a part of the impact of the MEES 2018 policy on the house prices should spill over to, from the rental market to the owner-occupied market, since all properties might be ultimately let out (i.e. a homeowner of an owner-occupied house might want to keep the option of renting the property out in the future). Hence, it is assumed that all properties with EPC rating F and G, independent of their type of tenure, are eligible to be affected in a similar way by the MEES 2018 policy.

[8] Intervention date is taken to be the 1st of April 2018

[9] In our analyses, we implicitly assume this measurement error to be random and non-systematic.

[10] For simplicity we only use the broadest classification into eight clusters, the so-called upper *Supergroup* tier. In addition to these 8 *Supergroups*, the classifications provides a set of more granular hierarchical clusters consisting of 26 *Groups*, and 76 *Subgroups*. We provide a description of each of these 8 *Supergroups*, or clusters, in Appendix A.

[11] The total number of 2011 OAs is 171,372 for England and 10,036 for Wales. `https://webarchive.nationalarchives.gov.uk/20160107193025/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/output-area--oas-/index.html`.

4.2 Creating the repeated sales data set

The original data set of property transactions lists 5,110,623 transactions between 2015 to 2019. In a first step, we construct the subsample of repeated sales, i.e. properties that were transacted both before and after the MEES 2018 policy intervention in April 2018. We then create a panel version of these data by identifying which transactions belong to the same property by using address information.[12] Creating this panel dimension allows us to examine how the price of the same property changes over time. This subsample includes 454,085 transactions. In a second step, we then match the energy performance certificates to each of these transactions using the common address. Specifically, we match each transaction with the closest prior registered EPC rating. If there was no prior EPC rating, the transaction we match it with is the EPC rating registered on the closest date afterwards.[13] This choice enables us to look at the full population of all relevant properties with EPC information, whilst also ensuring that the energy and house information is accurate and reflective of a property at time of sale. In the final step, we link each property in our matched data set with the corresponding geographic information using postcode information.

We also derive several additional variables from our matched data. First, we construct an indicator variable of whether a property is affected by the MEES 2018. We classify a property as being affected if it has an EPC rating of below E in its latest assessment. It takes on the value of one if that is the case and zero otherwise. Second, we create a variable which represents the high-level geographic region associated with the property. This variable is useful when examining regional effects. In this case, we adopt the *Nomenclature of Territorial Units for Statistics (NUTS)* Level 1 regions for the UK, and match the local-authority registered for each transaction with the corresponding region. The NUTS system is used for referencing the current administrative and electoral areas of the UK for statistical purposes. The NUTS Levels 1, 2, and 3 all stay fixed for a minimum of three years, with Level 1 referencing the regions, Level 2 the counties and grouped London boroughs, and Level 3 the unitary authorities and districts in the UK.

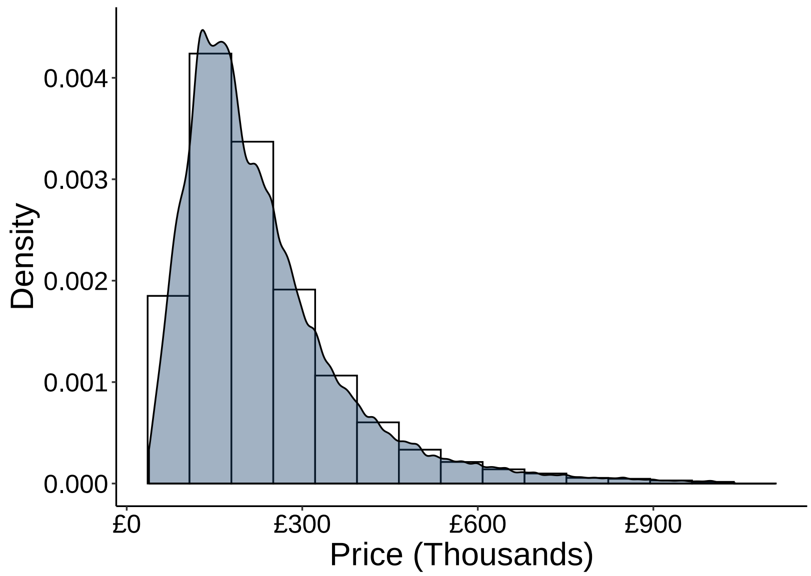
A simple analysis of the transaction data set shows some clear outliers. For instance, the minimum price is £1 and the maximum price is £247 million. As we wish to assess the impact on the bulk of the market and in order to avoid our results being driven by such outliers, we delete those observations that are below the 1st and above the 99th percentiles of this price distribution.[14] In the trimmed data set, there are 392,495 transaction. Their minimum price is £38,075 and their maximum is £1.04 million. We then remove all of those transactions with missing values in relevant variables, assuming the missing values are random. After ensuring that each property has at least one sale after the MEES 2018 policy intervention and one sale prior to this date, we are left with 305,337 transactions across 147,842 properties. In Figure 2, we show the empirical distribution of the property prices in our final data set. The distribution is right-skewed, i.e. a large share of properties are sold at prices between £150k and £250k with very few properties being sold at prices above £300k. The mean house price is £231k, with a standard deviation of £141k.

[12] Details on data processing can be found in Appendix A.

[13] This reflects the possibility that EPC might be registered only afterwards but could be available to prospective buyers at the time of the purchase.

[14] In robustness tests, we also trim the data set at other percentiles. Appendix C discusses the sensitivity of our results depending on the trimming.

Figure 2: Distribution of property prices



Note: This figure shows the empirical distribution of the price variable in our final data set, trimmed at the 1st and 99th percentiles. It shows a histogram of the price as well as a density plot of the price variable. Source: House price data are obtained from HM Land Registry.

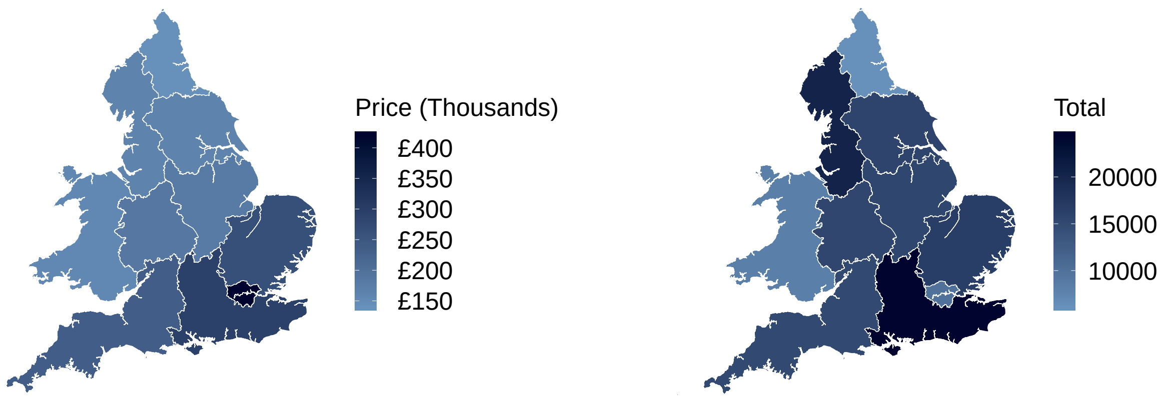
4.3 Descriptive analyses

Before moving on to our formal Difference-in-Difference analysis, we provide relevant descriptive statistics of our panel data set. Figure 3 presents the geographical distribution of properties in our sample. Panel a) shows the average transaction prices in each region. As expected, the highest average transaction prices are in London and in regions in the South-East of England. Panel b) illustrates the number of properties in each region. It shows a high concentration of properties in the South-East, with relatively few in Wales and the North-West.

Panel a) of Figure 4 shows the relative frequency of properties by EPC ratings before and after the MEES 2018 policy intervention. It illustrates several interesting stylised facts: First, the majority of properties are of medium energy efficiency with EPC ratings of C, D or E. Second, the share of properties with energy efficiency ratings of F and G is much lower than the share of properties with higher ratings. In fact, they make up only approximately 7% of the total sample size in the time period before the MEES 2018 policy intervention. In total, we observe 10,032 properties with such EPC ratings. This number is much lower than the 137,810 energy-efficient properties with EPC ratings of E or higher, which are not affected by the MEES 2018 policy intervention. Third, it appears that the share of properties with high EPC ratings increases over time, suggesting that properties become more energy-efficient after the MEES 2018 policy. We take this as evidence that a considerable number of properties appear to have upgraded their energy efficiency rating. From these descriptive analyses, it is unclear whether this is due to the MEES 2018 policy intervention, or other reasons for improving efficiency.

Figure 3: Prices and transactions by regions

(a) Average price in each region (b) Number of properties in each region



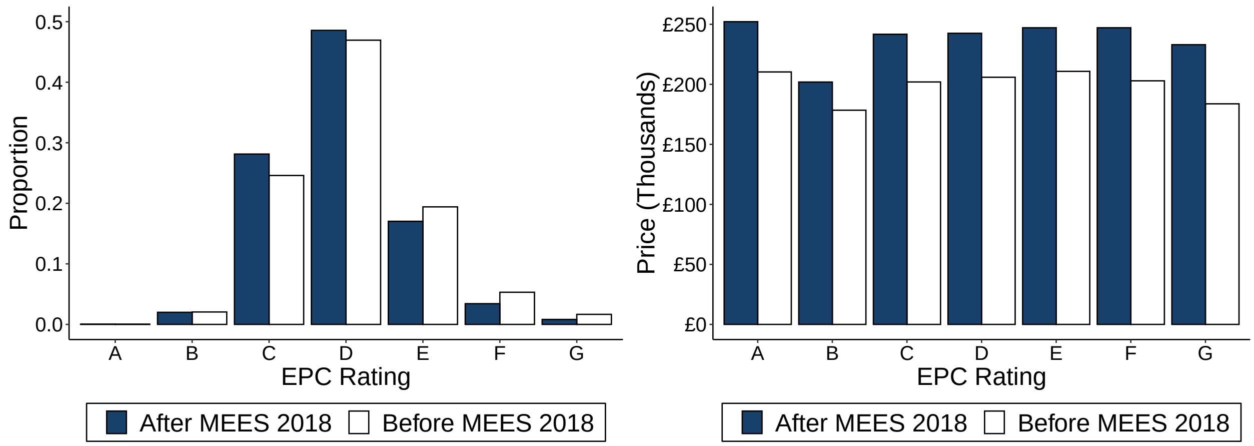
Note: This figure shows maps of geographic regions in England and Wales: Panel a) shows the average price in each region in our final sample of property transactions. Panel b) shows the number of properties in our final sample of property transactions. Source: House price data are obtained from HM Land Registry.

In Panel b) of Figure 4, we illustrate average transaction prices by EPC ratings over time, i.e. before and after the MEES 2018 policy intervention. We observe that the prices appear to rise across all EPC ratings. However, they seem to increase more for properties which are initially less energy-efficient. The descriptive evidence of these two panels of Figure 4 seems rather counter-intuitive. Over time, properties have become more energy-efficient, possibly suggesting higher demand for higher efficiency homes. By contrast, the prices in the least energy efficient class appear to rise the most. At this point, we must remind ourselves that these plots present a highly aggregated view and that there can be many confounding variables which impact price, besides the EPC rating. To truly understand the price fluctuations, and ultimately the impact of the MEES 2018 policy, we must understand how different energy-inefficient properties which are affected by the MEES 2018 are from energy-efficient properties, which are unaffected.

Figure 4: Number of properties and price by EPC over time

(a) Frequency of properties by EPC rating

(b) Average price by EPC rating



Note: Panel a) shows the relative frequency of EPC ratings before and after the MEES 2018 policy intervention. Panel b) shows barplots of the average transaction price by EPC ratings before and after the MEES 2018 policy intervention. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

4.4 Comparability of the treatment group and the control group

In this section, we examine the characteristics of properties in our sample. Table 1 compares characteristics depending on whether they are affected by the MEES 2018 policy intervention. On the one hand, there are those properties with EPC ratings below E. They are directly affected by the MEES 2018 policy intervention which applies to them. Hence, these properties form our *Treatment group*. On the other hand, there are properties with EPC ratings of at least E. They are not affected by the MEES 2018 policy intervention, our *Control group*. Just from examining these simple descriptive statistics, we can observe some sizeable differences. For instance, there are differences in the property type. As compared with more energy-efficient properties, those with EPC ratings below E are more often houses (77.4% vs 73.2%). Moreover, there is a higher share of older properties among them. Low EPC properties are more often built before 1900 than higher EPC ones (30.7% vs 11.3%), and in 1900-1966 (57.7% vs 43.2%). Similarly, the rural-residents demographic is more predominant in the low-energy efficiency properties (23.3% vs 9.3%).

Before explicitly accounting for these differences between properties in treatment and control groups, it is important to quantify the differences between these two groups. To assess these differences we calculate the *standardized mean difference (SMD)* for each covariate. The SMD value 0.1 is commonly used as a threshold for indicating imbalance (Stuart et al., 2013), suggesting that the treated and non-treated units are not comparable with respect to the corresponding variable.

In the case of continuous variables, such as TOTAL FLOOR AREA, the SMD is calculated as

$$\text{SMD} = \frac{\overline{X}_T - \overline{X}_C}{\sqrt{(S_T^2 + S_C^2)/2}} \,, \tag{2}$$

where $\overline{X}_T$ and $\overline{X}_C$ are the sample means of the variable for treated and non-treated properties, respectively, while S_T^2 and S_C^2 are the sample variances for the treated and non-treated properties.

In the case of multinomial variables SMD is calculated as

$$\text{SMD} = \sqrt{(T - C)^\top S^{-1}(T - C)} \tag{3}$$

where $T_j = P(X = \textit{category } j \,|\, \textit{treated})$, and $C_j = P(X = \textit{category } j \,|\, \textit{non-treated})$ for $j = 1, \ldots, k$, where X is the multinomial variable with k categories, S is the cross-category covariance matrix.

We report the calculated SMD values in column 2 of Table 1. As we can see, the two groups appear to be imbalanced in almost all covariates, with SMD values being greater than the critical threshold of 0.1. The most striking differences are in terms of the CONSTRUCTION AGE BAND, with a SMD value of 0.924, and the DEMOGRAPHIC characteristic reporting a SMD value of 0.397. The only exception is property TENURE whose SMD is equal to 0.071

Table 1: Characteristics of properties by EPC rating

Group by EPC rating	Treatment group	Control group	Control group
Sample	Full sample	Full sample	Matched sample
	(1)	(2)	(3)
PROPERTY TYPE		(SMD=0.203)	(SMD=0.017)
Bungalow	11.9%	9.4%	11.5%
Flat	9.3%	15.5%	9.1%
House	77.4%	73.2%	78.1%
Maisonette	1.4%	1.8%	1.3%
Park home	0.0%	0.0%	0.0%
CONSTRUCTION AGE		(SMD=0.924)	(SMD=0.008)
before 1900	30.7%	11.3%	31.0%
1900-1929	27.2%	15.6%	27.1%
1930-1966	30.5%	27.6%	30.3%
1967-1995	11.2%	30.1%	11.1%
1996-2006	0.4%	12.6%	0.4%
2007 onwards	0.0%	2.9%	0.0%
TENURE		(SMD=0.071)	(SMD=0.026)
owner-occupied	88.9%	86.6%	89.2%
rental (private)	10.4%	12.5%	9.9%
rental (social)	0.7%	0.9%	0.9%
DEMOGRAPHIC		(SMD=0.397)	(SMD=0.020)
Constrained city dwellers	5.0%	5.5%	5.0%
Cosmopolitans	4.1%	4.8%	3.9%
Ethnicity central	2.0%	2.7%	1.9%
Hard-pressed living	15.5%	17.9%	15.2%
Multicultural metropolitans	9.8%	9.9%	9.8%
Rural residents	23.3%	9.3%	23.9%
Suburbanites	16.1%	22.2%	15.9%
Urbanites	24.2%	27.7%	24.3%
REGION		(SMD=0.181)	(SMD=0.044)
East Midlands (England)	12.2%	10.4%	11.0%
East of England	10.6%	11.2%	11.0%
London	5.9%	6.8%	5.7%
North East (England)	2.8%	4.3%	2.7%
North West (England)	13.5%	13.8%	13.4%
South East (England)	13.2%	16.7%	13.3%
South West (England)	12.7%	10.0%	13.4%
Wales	7.6%	5.5%	8.0%
West Midlands (England)	11.0%	10.4%	10.8%
Yorkshire and The Humber	10.4%	10.7%	10.8%
TOTAL FLOOR AREA		(SMD=0.160)	(SMD=0.016)
Floor area	$90.69m^2$	$84.47m^2$	$91.38m^2$
HABITABLE ROOMS		(SMD=0.155)	(SMD=0.007)
less than 4	22.3%	28.8%	22.4%
4 to 5	58.2%	54.5%	57.8%
more than 5	19.5%	16.7%	19.7%
Number of observations	(N=10,032)	(N=137,810)	(N=10,032)

Note: This table examines the balance of each covariate: Column 1 shows covariates of properties with an EPC rating below E, our *Treatment group*. Column 2 shows covariates of properties with an EPC rating of at least E, our *Control group*, in the full sample. Column 3 shows covariates of properties with an EPC rating of at least E, our *Control group*, in our matched sample after PSM matching. For each categorical covariate, percentages of each level are reported. For the continuous covariate TOTAL FLOOR AREA, the mean is displayed. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG. Demographic data are from the ONS.

This imbalance between our treatment and control groups raises the concern that comparing transaction prices across these two groups might suffer from a selection bias. For that reason, we seek to ensure comparability between properties in these two groups by employing *Propensity score matching* (PSM) (Rosenbaum and Rubin, 1983). In our set-up, PSM aims to match each property that is assigned to the treatment group with one property assigned to the control group. This ensures that the properties we examine in our empirical analyses are similar on observable characteristics, such as property size, region or property age. PSM can be used to reduce selection bias that could influence the estimation of the effects of the MEES 2018. Specifically, we implement PSM on the properties before the MEES 2018 policy intervention. If a property was sold multiple times before the MEES 2018 policy, we match them based on the transaction whose date was closest to 1 April 2018, the day of the MEES 2018 policy intervention.

We conduct the *Propensity score matching* (PSM) in two steps, see Appendix B for details. In a first step, we estimate the propensity of being affected by the MEES 2018 policy intervention, conditional on a set of observed covariates (Rosenbaum and Rubin, 1983; Rubin, 2007). The "propensity score" is the predicted probability of the treatments assignment. In a second step, we then use these propensity scores to find for each "treated" property the most similar property unaffected by the MEES 2018. For properties with similar propensity scores, the idea of PSM is that the treatment assignment for these two properties is independent of all confounding variables (Westreich et al., 2010), i.e. the selection bias decreases.

To estimate propensity scores, the existing literature had typically used one of two parametric models, either Logistic regression or Probit regression models (Caliendo and Kopeinig, 2005). More recently, there has been an increasing use of machine-learning methods to construct scores (Westreich et al., 2010; Cannas and Arpino, 2019; Luellen et al., 2005; Setoguchi et al., 2008; McCaffrey et al., 2004). For example, Lee et al. (2010) show that machine-learning methods can improve the covariate balance between treated and non-treated units in the resulting matched data set using a simulation study. For that reason, we complement a standard Logistic regression with two popular tree based machine-learning classifiers. Since single tree classifiers have the unfortunate reputation of being rather weak prediction models (James et al., 2013, p. 316; Efron and Hastie, 2016, p. 324), we employ the following two ensemble methods that usually enjoy good predictive performance:

Random Forest is a method that averages across many trees based on a bootstrap resampled data set. Further to simply bagging the trees, the method attempts to decorrelate the trees by further resampling of the covariate set (over which each split is estimated) (Breiman, 2001). For each tree a random sample of $m < p$ covariates are chosen for inclusion. A typical value of this hyperparameter is $m = \sqrt{p}$, however, this can also be assessed via cross-validation. Using random forest with $m = p$ is equivalent to a bagged tree model (Efron and Hastie, 2016, p. 327).

Boosted Trees work by repeatedly extending a basic seed model (Friedman et al., 2000). More specifically, for each iteration of the algorithm, a shallow classification tree is added to the previous model (estimated based on the residuals), and hence we build up an additive model consisting as a sum of trees. Overall, boosting has three hyperparameters that can alter the predictive performance of a boosted model: the number of trees B, the number of splits in each tree d, and the shrinkage parameter (which scales each tree to be added) λ (Efron and Hastie, 2016, p. 333-334).

When fitting the above models, we split the data into training and test sets, respectively 70 and 30 percent of the data, and we also ensured the class balance between treatment and control groups was consistent across these sets. Logistic regression models require no tuning parameters and we use a full model with all covariates. However, for random forest and boosting, we implement 5-fold cross-validation (CV) (Hastie et al., 2009, p. 241-249) on the training data. The particular implementation of boosting used was via the *eXtreme Gradient Boosting (XGBoost)* package (Chen and Guestrin, 2016)[15].

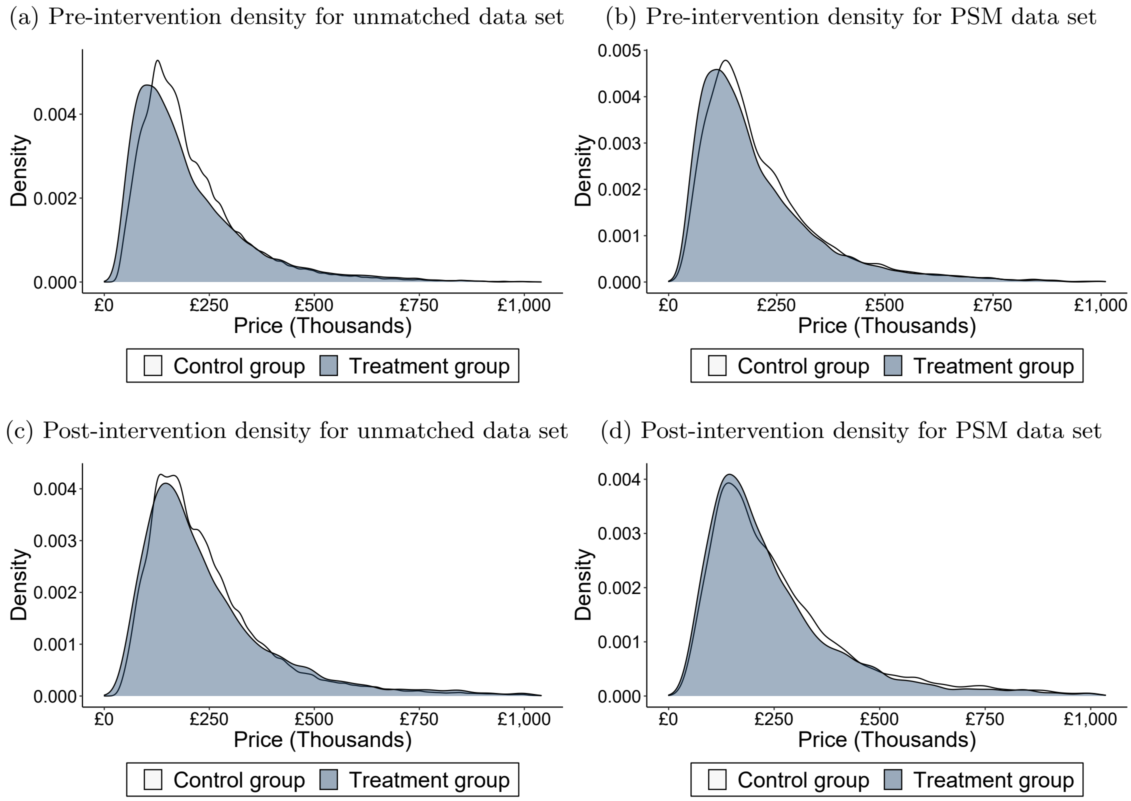
For each model, the hyperparameters that maximized the cross-validated *area under the curve (AUC)* score were selected. One way to interpret the AUC score, is to think that if we randomly choose a positive instance and a negative instance, then AUC represents the probability that the classifier ranks the positive instance higher than the negative instance. For the random forest, the hyperparameter set of values that maximized the CV AUC score was equal to $m = 2$, while for XGBoost was 500 trees, $d = 6$ and $\lambda = 0.01$. The final performance of each method in terms of AUC is given in Appendix B. Interestingly, in this application all methods perform similarly by this metric, with a slight edge given to the boosting method (XGBoost). In addition to the slightly higher AUC, it is now clear the boosted method provides superior alignment of covariate balance as measured through SMD. For this reason, we choose the boosted model to continue our analysis confident that the sample is far more comparable than the unmatched data set.

Indeed, PSM matching improves the covariate balancing across treatment and control groups. In column 3 of Table 1 we now show covariates and SMD values of each coviariates of our matched data set, using the Boosted model. We can see that SMD values are now lower than 0.1 for all variables. Some key examples of this can be observed by comparing column 1 and column 3. For instance, after matching the mean total floor area for treated properties was only 0.96m^2 smaller than for non-treated properties (as opposed to 5.3m^2 in column 2), while the percent difference between treated and non-treated properties built before 1900, and in the periods 1967-1995 and 1996-2006, was reduced to -0.2% from 20%, 0.1% from -19.1%, and 0% from -12.8%, respectively. We conclude that any difference that existed between the treatment and control group prior to matching has been reduced dramatically after implementing PSM.

We also examine how the price distributions of properties in our treatment and control groups change following the implementation of PSM. In Figure 5, we plot density estimates of transaction prices both before ("pre") and after ("post") the MEES 2018 policy intervention. Panel a) shows the distributions of prices in our unmatched data set before the MEES 2018 policy intervention. First, it shows prices of properties with EPC ratings of at least E, our control group. It also illustrates the distribution of prices of those properties with EPC ratings of below E, our treatment group, indicated by the shaded area. These prices are on average lower than prices of energy-efficient properties, indicated by the white area under the curve. Panel b) then illustrates the distribution of prices in our matched sample where we keep only those energy-efficient

[15]The primary benefit of this package is that it enables the usage of parallel computing to boost the speed over the classical boosting algorithm, whilst also including some options for additional model regularisation.

Figure 5: Distribution of transaction prices of properties



Note: This figure shows density estimates of transaction prices by energy efficiency. Our control group consists of properties with EPC ratings of at least E. Our treatment group consists of properties with EPC ratings below E. Plots in the top row show density estimates before ("pre") the MEES 2018 policy intervention and those in the bottom row show density estimates after ("post") the MEES 2018 policy intervention. Plots on the left show densities of the unmatched data set before implementing PSM and those on the right show the densities of the matched data set after PSM matching. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

properties that are similar to our energy-inefficient properties. In this matched sample, the distribution of prices of energy-inefficient properties shifts to the left. I.e. these properties in our control group not only become more similar in terms of observable characteristics, but also their price distribution becomes more similar to the distribution of prices of energy-inefficient properties. In the bottom part of this figure, we then illustrate the distributions of prices after MEES 2018 policy interventions. Comparing Panel c) with Panel d), we observe, also for this time period, that properties become more similar in terms of their price distribution following the implementation of PSM.

5 Main Analyses

In this section, we estimate the effect of the MEES 2018 policy intervention on transaction prices from a sample of repeated property transactions. Ideally, we would like to estimate the effect of this policy by calculating the difference in the average transaction price of energy-inefficient properties that have been affected by the policy (*Treatment group*) and the average hypothetical, counterfactual transaction price of the same properties if they had not been affected by the policy (*Control group*).[16]

Unfortunately, we cannot directly observe such counterfactual outcomes. *Randomized controlled trials (RCT)* can address this problem. In such empirical designs, properties would have the same chance of being allocated to the treatment group, i.e. being affected by the MEES 2018 policy intervention, or to the control group, i.e. not being affected by it. RCTs ensure that both groups are equivalent for comparison, and that there are not any pre-existing differences. In such a setting, we would not expect any confounding variables, e.g. neither the property type, size, age, nor the location of a property would influence the outcome and confound the results of the analysis. However, it is easy to see that in case of the MEES 2018 policy intervention, the treatment assignment is not random. The least energy-efficient properties, meaning those with EPC ratings of F or G, are automatically allocated to the treatment group.

Since we cannot assume such a random allocation, we adopt a quasi-experimental design to estimate the effect of the policy intervention on transaction prices. A quasi-experiment is a type of observational study, that is similar to an RCT, but without randomization (Dinardo, 2010), i.e. the allocation of units, in our case properties, to the treatment and control groups is not randomized. Instead, it depends on pre-specified criteria that does not adhere to randomness (e.g. a threshold value). The MEES 2018 offers a clear selection criterion since the least energy efficient properties in England and Wales, meaning those rated F or G on their EPC, are automatically allocated to the treatment group, while the remaining ones are in the control group.

In our empirical analyses, we exploit this criterion. Specifically, we employ a Difference-in-Difference design which has been widely used in the applied literature (e.g. Card and Krueger, 1994). It estimates the effect of the MEES 2018 policy intervention on transaction prices by calculating the average change in transaction prices among properties in the treatment group, i.e. those properties affected by the MEES 2018. It then compares it to the average change in transaction prices among properties in the control group, i.e. properties unaffected by the MEES 2018 policy intervention.

The following subsections detail our Difference-in-Difference analysis, alongside the measures we take to verify and ensure the validity of its underlying assumptions. To verify the parallel trends assumption, we propose to use a *Generalised Additive Model (GAM)* construction. The use of GAMs is novel in the existing literature and we offer a discussion on why this method might be superior to existing approaches also describing the basic mechanics of this method.

[16]Or in more technical terms, the effect of the MEES 2018 policy intervention on the outcome variable, our case the transaction price, can be then estimated by comparing the identified counterfactual outcomes to the observed outcomes under the policy intervention (Fougère and Jacquemet, 2020).

5.1 Testing the parallel trends assumption

The inference for the estimated intervention effect using our Difference-in-Difference (DiD) design is valid only under the assumption that the outcomes in treatment and control group follow the same time trend in the absence of the policy intervention (Lechner, 2010; Ryan et al., 2019). In the DiD setting this is known as the *parallel trend assumption* or *common trend assumption*. Clearly, we cannot verify this counterfactual. However, we can look for evidence against the assumption in the time leading up to the MEES 2018 policy intervention. Standard approaches to verifying (or not rejecting) this assumption do not typically go beyond visual inspections of trends in the outcome variable, or testing for differences in linear pre-intervention trends (Ryan et al., 2019; Roth, 2020; Rambachan and Roth, 2019). Given the importance of this assumption, we argue that a more realistic approach is to avoid the assumption of a linear relationship between outcome and time. Instead, we propose the observed data to determine the shape of the trend in outcomes. To that end, we present a novel use of *Generalized Additive Models (GAMs)* (Hastie and Tibshirani, 1986). The advantage of a GAM model is that it allows for smooth non-linear trends over time.

Overview of GAMs

The basic GAM model enables non-linear functions of some, or all, of the covariates, while maintaining a simple additive model structure (Hastie and Tibshirani, 1986). Specifically, GAMs have the following form:

$$y_i = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}) + \epsilon_i, \quad i = 1, 2, ..., n \tag{4}$$

where y_i is the ith observation of outcome Y, x_{ij} is the ith observation of explanatory variable X_j, and f_j is a non-linear function for explanatory variable X_j. GAMs have been used to model trends as a smooth, non-linear function over time for a wide variety of data, such as for palaeoenvironmental time series (Simpson, 2018), high-frequency water-quality data Yang and Moyer (2020), or survey data for bird populations (Fewster et al., 2000). In this paper, we set up the GAM to use factor-smooth interactions, which allows us to formally test for evidence against the parallel trends assumption. Before implementing the GAM on the PSM-derived matched data set of comparable pre-intervention properties, we first discuss the underlying structure of the generalized additive model approach.

In order to move away from imposing a predetermined linear form on f_j, the GAM attempts to estimate the shape of f_j from the data whilst imposing smoothness constraints to avoid overfitting. Given these restrictions, the f_j non-linear functions in equation (4) are referred to as *smooth functions (smooths)*. In order for the smooth to be represented in a parametric form that can be estimated, it needs to be specified using a set of *basis functions*. Let $b_{kj}(X_j)$ represent the kth basis function for the smooth of covariate X_j. Assuming that for X_j there are K basis functions, then the smooth $f_j(X_j)$ can be represented as:

$$f_j(X_j) = \sum_{k=1}^{K} \beta_{kj} b_{kj}(X_j) \ , \tag{5}$$

where β_{kj} is the coefficient of the kth basis function that must be estimated (Wood, 2017, p. 162).

There are many ways to specify basis functions, such as in cubic polynomial regression, where the basis functions are: $b_{1j}(X_j) = X_j$, $b_{2j}(X_j) = X_j^2$, and $b_{3j}(X_j) = X_j^3$. We investigate a number of different basis functions for fitting the GAMs in this paper (e.g. thin plate regression splines, cubic regression splines, P-splines). They all give the same conclusions. To keep the paper concise, we only give details for the piecewise cubic polynomial model, under the constraint that the first and second derivatives of the piecewise polynomials are continuous at the knots, i.e. that the piecewise polynomial must be continuous and smooth. The parameters of the GAM model are traditionally estimated via a *backfitting algorithm* (Hastie and Tibshirani, 1986).[17]

Testing against the parallel trend assumption

Looking at equation (4), it is easy to see that using the Price variable as the outcome we can then use a GAM to model trend as a smooth non-linear function of a time covariate. Yet, in order to see whether treated and non-treated properties follow a similar house price trend over time (prior to the MEES 2018 policy intervention date) modelling the overall price trend is irrelevant. What we truly desire, is to understand trends over time, for the different levels of the binary variable $I(\mathrm{EPC_i} \geq \mathrm{E})$, which indicates whether a property is in the *Treatment group* or in the *Control group*. This can be achieved by fitting a GAM with factor-smooth interactions, which estimates a separate smooth trend for each level of a factor. Thus, one can examine whether the smooth trends of the two levels are different. This approach was implemented by Rose et al. (2012) to estimate trends in the pollutant profiles at individual sites in three regions of Scotland. However, the application to testing the assumptions for the DiD procedure appears novel.

Specifically, our implementation uses the following model

$$\mathrm{Price}_i = \beta_0 + \textstyle\sum_{j=1}^{p} \alpha_j X_{ji} + f(\mathrm{Time}_i) + (\beta_\Delta + f_\Delta(\mathrm{Time}_i))\, I(\mathrm{EPC_i} \geq \mathrm{E}) + \epsilon_i \ , \tag{6}$$

where $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, ..., n$, and $I(\mathrm{EPC_i} \geq \mathrm{E})$ denotes an indicator function, if the property is in the treatment group then this function is zero, otherwise it is one. The α_j terms represent $j = 1, \ldots, p$ fixed effects based on property characteristics X_{ji}, while Time is a continuous time covariate.[18] As for the model parameters, β_0 is the intercept that represents the mean house price of a treated property, β_Δ is the difference between the mean price of a non-treated property and a treated property. The smooth functions $f(t)$ and $f_\Delta(t)$ respectively represent the price trend for the treatment group, and the difference in trends (between treated and non-treated groups). The GAM was implemented using cubic regression spline smooths, with the knots distributed evenly throughout the time covariate values.

We used the *mgcv* package (Wood, 2017) in R to estimate the GAM model parameters, and at this stage

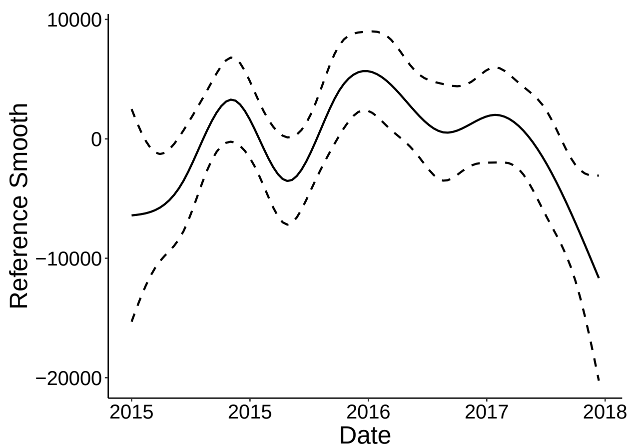
[17] In this paper, we use the *mgcv* R package following guidance provided in the text book by Wood (2017)

[18] In our construction time is represented as the number of days from the first transaction in our sample. We control for a large set of characteristics that may be relevant for hedonic pricing models, i.e. X={REGION, DEMOGRAPHIC, PROPERTY TYPE, TOTAL FLOOR AREA, CONSTRUCTION AGE, TENURE, HABITABLE ROOMS}.

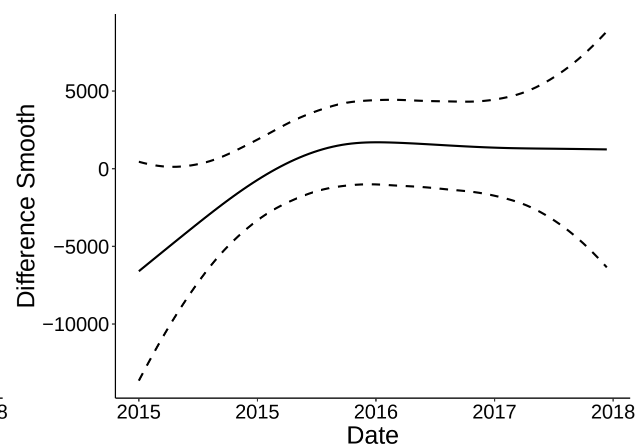
our focus is on the inference for the difference smooth $\hat{f}_{\Delta}(t)$. Using our construction the F-statistic reported for this smooth is calculated under the null hypothesis: H_0 the difference between the smooth trends of treated and non-treated properties is equal to zero, with the alternative H_1 being that it is different from zero. An F-statistic that corresponds to a small p-value suggests that the trends for the two groups are different, in our case, we obtain a p-value of 0.234 (F-statistic 1.617) for the significance of $\hat{f}_{\Delta}(t)$, and a p-value of 0.001 (F-statistic 3.385) for the reference smooth $\hat{f}(t)$.[19] The estimated functions $\hat{f}(t)$ and $\hat{f}_{\Delta}(t)$ are plotted in Figure 6. In this case, it is clear the estimated price trend for the treatment group is non-linear (Fig. 6a), however, the difference smooth $\hat{f}_{\Delta}(t)$ (Fig. 6b) possesses a confidence interval that consistently includes zero across time. As such, when coupled with the summary F-statistic we find there is insufficient evidence to reject the null hypothesis at the 5%, or even 10% level. Whilst we should never use lack of evidence against a null to accept it, in this case, the parallel trends assumption does seem reasonable and gives us confidence to proceed with our Difference-in-Difference (DiD) analysis.[20]

Figure 6: Fitted price trends

(a) Fitted price trend for treated properties



(b) Estimated smooth for the difference



Note: Panel a) shows the fitted price trend for treated properties $\hat{f}(t)$ from the GAM of Eq. 6. Panel b) shows the estimated smooth for the difference between treatment and control groups $\hat{f}_{\Delta}(t)$. Dashed lines represent the estimated 95% confidence interval. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

[19]It is worth noting that there is a difference in property prices based on treatment/control ($\hat{\beta}_{\Delta}$ =10,500, p-value < 0.001), however, such constant differences are allowed under the parallel trend assumption.

[20]To verify the robustness of our conclusions we also estimated a GAM model which included reference and difference smooths for each region, e.g. London, North-East, etc. While different regions possessed different reference functions, the difference smooths were never significantly different from zero, thus the parallel trends assumption appears reasonable even at a regional scale.

5.2 Difference-in-Difference analyses

In our Difference-in-Difference (DiD) design, we quantify the effect of the MEES 2018 policy intervention by estimating the change of transaction prices of properties affected by it relative to properties that are unaffected by it. For this estimation, we use our matched sample consisting of properties that are very similar on observable characteristics. Specifically, we examine the subset of $n = 20,064$ properties with at least two transactions, one before and one after the policy intervention on 1 April 2018.[21]

In this sample, we observe each property over time whenever it is transacted. This panel dimension allows us to employ property fixed effects in our regression model. Making use of the latter enables us to estimate the effect of the MEES 2018 *within* a property over time, thus removing the effects of unobserved time-invariant confounding variables *between* the properties. Such confounders may include the size and the architectural style of the building and features of the location of the property such as its proximity to the sea.[22] Specifically, we estimate the following regression model in order to provide inference for the DiD effect:

$$\text{Price}_{it} = c_i + \delta\ I(\mathrm{Post_t}\ \times\ \mathrm{Treated_i}) + \tau\ I(\text{Post}_t) + u_{it}\ , \tag{7}$$

where the subscript $i = 1, \ldots, n$ indexes the properties, and $t = 1, 2$ denotes the pre and post intervention periods. An interpretation of the model components is given below:

- $I(\mathrm{Post_t}\ \times\ \mathrm{Treated_i}) = 1$, if property i has an EPC rating below E at time $t = 2$, meaning for a treated property in post-intervention period, and zero otherwise;
- $I(\text{Post}_t) = 1$, if $t = 2$, i.e. the transaction date is post-intervention, and 0 otherwise;
- c_i is the intercept for the ith property, that controls for unobserved variation at a property level;
- u_{it} are idiosyncratic errors;

Of primary interest is the estimated coefficient of $Post\ \times\ Treated$ which measures the change of the transaction price of properties affected by the MEES 2018 policy relative to unaffected properties, those with EPC ratings of at least E.[23]

Table 2 summarises our results. Column 1 shows the estimated coefficients using the full sample of all properties in our matched data set. The estimated coefficient of $Post\ \times\ Treated$ points to a decrease in the prices of affected properties with EPC ratings below E by £9,200 relative to unaffected properties with an EPC ratings of at least E. This estimate is seen to be statistically significant at the 1% level of significance. In column 2, we show the same analysis but on a subsample of properties that were deemed very likely to be rented. Specifically, we focus on all those properties that have been privately rented at any point before or after 1 April 2018. This resulted in a subsample of 5,002 transactions from the matched data set. The estimated effect appears slightly smaller in this subsample of rented properties, decreasing to about £7,150.

[21] To create a balanced panel, we only keep the post-intervention observation with the most recent transaction date.

[22] We report the estimated *within* R^2 which measures the proportion of the variation in the Price variable within properties, that is explained by the respective model.

[23] We fit the regression model (7) in *R* using the *plm* package (Croissant and Millo, 2008).

This contrasts our theoretical priors, as we expected that the policy affected predominantly existing landlords of private-rented properties. However, the magnitude of this point estimate does not appear to be very different from the full sample after one considers estimation uncertainty. Indeed, standard errors, which we report in brackets, are almost three times as large in this subsample compared to the full sample. Overall, the MEES 2018 policy intervention appears to have significantly impacted the house prices of affected properties, both in the full population of all properties and in the sub-population of properties that are rented.

Table 2: Effect of the MEES 2018 policy intervention on house prices

Sample	Full Sample	Private Rented
Dependent variable	Price	Price
	(1)	(2)
Post $\times$ *Treated*	$-9{,}200$	$-7{,}150$
	(1,060)	(2,930)
	$p < 0.0001$	$p = 0.015$
Post	51,500	45,000
	(585)	(1,530)
	$p < 0.0001$	$p < 0.0001$
Property F.E.	YES	YES
Observations	40,128	5,002
R^2 (*within*)	0.333	0.304

Note: This table shows the results of the regression model in eq. (7) on both the full balanced panel in column (1), and the subsample of properties that are private rented in column (2). Standard errors are reported in brackets. All specifications include property fixed effects. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

We conduct several sensitivity tests. First, in our analysis the temporal dimension may cause the idiosyncratic errors u_{it} to be correlated within properties. This type of correlation can lead to significantly smaller standard errors, narrow confidence intervals, large t-statistics, and small p-values (Cameron and Miller, 2015). In order to address this concern, we cluster standard errors by individual property to allow correlation to exist in the u_{it} errors for the same ith property (Cameron and Miller, 2015, p. 318). For completeness, we present the adjusted parameter estimates in Appendix C (Table 9). The resulting p-values are only slightly altered, with the effect size being exactly the same. Hence, these results indicate that auto-correlation is not a large issue. Second, we further assess the sensitivities of our results with respect to outliers. Specifically, we report the results of an analysis that trims our data at the 5th and 95th percentile of the Price variable. These results are presented in Appendix C (Table 10). Interestingly, the size of the estimated effects of *Post* $\times$ *Treated* decreases slightly from £9,200 to £5,040 (full sample) and £7,150 to £4,700 (private rented sample), respectively. Both estimates remain statistically significant at the 5 percent level.

Overall, we conclude that prices of the least energy-efficient properties affected by the MEES 2018 policy intervention decreased by about £5k to £9k relative to unaffected ones, with £9k being our best estimate.

6 Discussion and Conclusion

Our results suggest that prices of properties affected by the MEES 2018 policy intervention decreased by about £5k to £9k relative to unaffected ones. In this section, we offer a discussion which aims to put the size of this estimate into perspective. In particular, we discuss some potential implications of the policy for financial stability and the wealth distribution among homeowners in the UK. Finally, we discuss how our results may generalise to the full population of all properties.

6.1 Implications of the MEES 2018 policy for financial stability

Decreases in property prices mean that collateral values of outstanding mortgages decrease. Once outstanding loan amounts exceed the value of collateral, mortgage lenders would incur losses if borrowers defaulted on their mortgage payments. This can be a contributing factor for financial instability depending on the share of such mortgages outstanding on banks' balance sheets. In this section we examine this concern using a data set employed by Guin and Korhonen (2020). It includes mortgages outstanding at year-end 2017, i.e. right before the MEES 2018 policy intervention was introduced. Using these data, we calculate the distribution of the so-called loan-to-value (LTV) ratio of these outstanding mortgages. The LTV ratio measures the outstanding loan amount relative to the property value of the last mortgage transaction. In our sample, the median LTV ratio is about 70%. The share of mortgages with LTV ratios above 90% is about 17%.

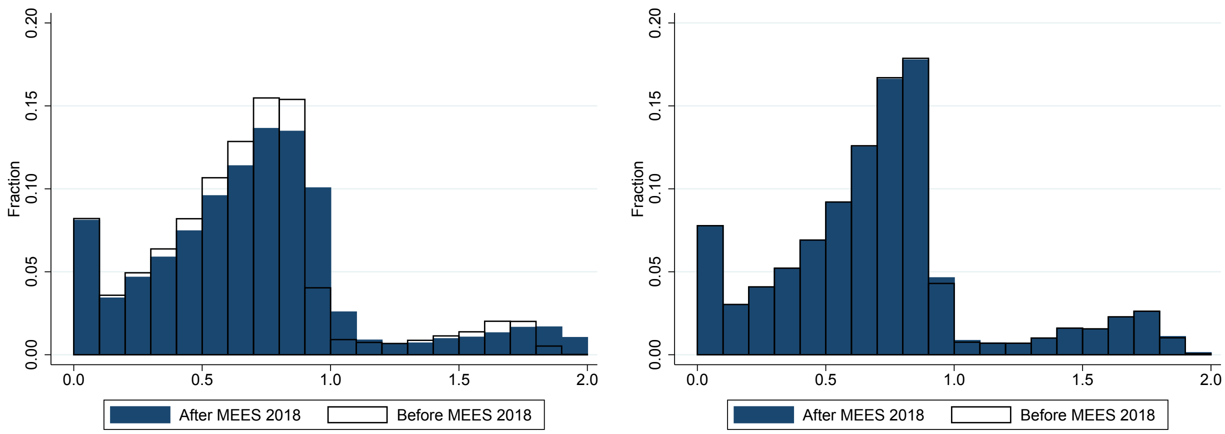
The MEES 2018 policy intervention affected only mortgages against the least energy-efficient properties with EPC ratings of F or G. Panel A of Figure 7 shows their distribution as indicated by the white bars. In our sample, their median LTV ratio is about 66%. The share of mortgages with LTV ratios above 90% is about 14%, which indicates that LTVs of these mortgages were slightly lower than in the full sample of all outstanding mortgages.

To gauge the effect of the MEES 2018 policy intervention on the distribution of LTV, we conduct a back-of-the envelope calculation. Specifically, we subtract the estimated effect of the MEES 2018, around £9k, from the value of each property that was affected by the MEES 2018, all those properties with EPC ratings of F or G. The navy bars in Panel A of Figure 7 show how the distribution of these affected properties after the MEES 2018 policy. It is visible that the MEES 2018 shifts the LTV distribution to the right. In particular, the share of mortgages with LTV ratios between 90% and 100% increases substantially.

However, mortgages against energy-inefficient properties affected by the MEES 2018 policy intervention make up only a small fraction of all mortgages. Panel B of Figure 7 shows the LTV distribution of all mortgages. It is difficult to visually detect a change in the distribution following the MEES 2018. Our back-of-the envelope calculation suggests that the MEES 2018 pushed only about 0.5% of mortgages outstanding in end 2017 into the range of outstanding loan-to-value ratios of at least 90%. The share of mortgages that had already been in this range of at least 90% LTV increases by 2.8%. We take this as evidence that adverse implications for financial stability are limited.

Figure 7: LTV distribution of mortgages outstanding in 2017

(a) Mortgages against energy-inefficient properties (b) All mortgages



Note: This figure shows the LTV distribution of mortgages outstanding year-end 2017. Panel A shows the distribution of mortgages against properties with EPC ratings below E. Panel B shows the distribution of all mortgages. Data source: Guin and Korhonen (2020).

6.2 Implications of the MEES 2018 policy for wealth inequality

Our results suggest that the MEES 2018 policy intervention has decreased the prices of the least energy-efficient properties relative to more energy-efficient ones. This might have implications for the wealth distributions among home owners if property values are not evenly distributed across energy efficiency ratings. Indeed, the least energy-efficient properties are less expensive than more energy-efficient ones, as indicated in Figure 4. The MEES 2018 should have decreased the price of these properties by about £9k which meant the wealth inequality among UK homeowners has decreased.

To quantify the magnitude of wealth inequality among UK homeowners, we calculate the Gini coefficient. It is a widely established measure of statistical dispersion intended to represent the income inequality or wealth inequality within a country or any other group of people (Fagereng et al., 2016). A Gini coefficient of zero expresses perfect equality, where all values are the same. By contrast, a Gini coefficient of one (or 100%) indicates maximal inequality.

For the purpose of this analysis, we assume that each homeowner owns one property as we cannot identify homeowners with multiple homes in the data. Using our sample of repeated property transactions, we estimate a Gini coefficient of about 34.40% before the MEES 2018 (first row of Table 3). The order of magnitude of this estimate compares well with previous estimates for UK homeowners (Levin and Price, 2010). After the MEES 2018, the Gini coefficient of the same properties, and hence homeowners, drops to about 33.40%, reflecting changes in house prices which occurred over time (second row of Table 3). This means that the distribution of housing wealth has become more equal over time despite the implementation of the MEES 2018 policy intervention.

But how would the distribution of housing wealth have evolved if the MEES 2018 policy had not been implemented? To gauge the effect of the MEES 2018 on the wealth distribution of UK homeowners, we then construct this counterfactual. Specifically, we examine the Gini coefficient if the MEES 2018 had not been implemented. We construct it by adding the estimated effect of the MEES 2018, i.e. about £9k, to the values of the least energy efficient properties being affected by this policy intervention. This counterfactual analysis suggests that, in absence of the MEES 2018, the Gini coefficient would have decreased by another 1 percentage point further, more precisely decreasing to 32.23% (third row of Table 3). We take this as evidence that the MEES 2018 policy intervention might have led to moderate increases in the wealth inequality among UK homeowners. In terms of magnitude, this one percentage point decrease in the Gini coefficient due to MEES 2018 compares well to the natural change in the Gini coefficient in the time period of 2015-2019 which we consider.

Table 3: Gini coefficients before and after the MEES 2018 policy intervention

	Gini coefficient	Lower bound	Upper bound
	(1)	(2)	(3)
Before MEES 2018	34.40%	34.08%	34.72%
After MEES 2018	33.40%	33.08%	33.70%
After MEES 2018 + £9,000 (treatment)	32.23%	31.67%	32.76%

Note: This table presents estimates of the Gini coefficient of house prices in the balanced panel data set of 40,128 transactions. The first row shows the Gini coefficient before the MEES 2018. The second row shows the Gini coefficient after the MEES 2018. The third row shows the Gini coefficient after the MEES 2018 adding the estimated effect of £9k to properties with low EPC ratings. Column 1 shows the estimated Gini coefficient. Columns 2 and 3 show upper and lower bounds of the non-parametric 95% bootstrap confidence intervals, respectively. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

6.3 External validity

While the DiD methodology implemented here in conjunction with PSM provides a powerful tool to reduce the impact of confounders on causal inference, there may be limitations to the external validity of our results. In this section, we offer a brief discussion on how our analysis may generalise to the rest of the UK housing market. If one considers where biases may enter our analyses, an obvious point to consider is when we construct our sample selection, which focuses on properties with repeated sales prior and post intervention.

To address this concern, we compare the characteristics of properties in our sample of repeated sales properties with a random sample of all properties in England and Wales, which should be representative of the full population. We present this descriptive comparison in Table 4, where column 1 shows the characteristics of the random sample. By contrast, column 2 illustrates the characteristics of our repeated sales sample employed in our empirical analyses. For many variables, we see close alignment, for instance the size of properties, as measured by TOTAL FLOOR AREA, and CONSTRUCTION AGE. However, REGION, DEMOGRAPHIC, and most of all TENURE, present significant differences. The most striking of these is the tenure of the properties. In the random sample of the full population of all properties, we observe that 19% of properties are socially rented. However, these properties are all but absent from our repeated-transaction sub-sample. Staying with the TENURE variable, we see that there is also a large imbalance with private rentals in our data set (12.4%) compared to the random sample (21.8%). It appears that properties that are likely to be rented are transacted less often either side of the intervention period. Digging into this pattern further, we note that our repeated sub-sample contains a relative lack of properties within London. This suggests that properties in the capital were not sold as frequently, an attribute which may be correlated with either the presence of more social housing, and/or privately rented houses being transacted less in this region. Regarding the demographic variables, the imbalance in these is largely expected given the regional discrepancies, specifically our relative under-sampling of London properties.

Comparing the repeatedly transacted properties with the random sample suggests our results may not tell the whole picture, especially in the London region. Our results should also not be projected onto the value of social housing stock, which due to the requirement for repeated sales is very hard to assess empirically via market pricing. The relative lack of private rented housing in our sample may bias the estimated effect in the likely to be rented sub category, where we saw the intervention had a slightly smaller DiD effect size. However, with these discrepancies (which are primarily London centric) in mind, the repeated sub-sample seems reasonably representative for the other regions of England and Wales.

Table 4: Comparison of property characteristics (population vs. repeated sales)

Sample	Population (random sample)	Repeated sales
	(1)	(2)
PROPERTY TYPE		(SMD=0.272)
Bungalow	9.4%	9.6%
Flat	25.3%	15.1%
House	62.7%	73.5%
Maisonette	2.6%	1.8%
Park home	0.0%	0.0%
CONSTRUCTION AGE		(SMD=0.106)
before 1900	11.7%	12.6%
1900-1929	15.2%	16.4%
1930-1966	31.7%	27.8%
1967-1995	29.4%	28.8%
1996-2006	9.6%	11.7%
2007 onwards	2.5%	2.7%
TENURE		(SMD=0.749)
owner-occupied	59.0%	86.8%
rental (private)	21.8%	12.4%
rental (social)	19.2%	0.9%
DEMOGRAPHIC		(SMD=0.380)
Constrained city dwellers	9.0%	5.5%
Cosmopolitans	7.0%	4.7%
Ethnicity central	7.6%	2.7%
Hard-pressed living	17.7%	17.8%
Multicultural metropolitans	14.2%	9.9%
Rural residents	9.8%	10.3%
Suburbanites	15.9%	21.7%
Urbanites	18.8%	27.5%
REGION		(SMD=0.287)
East Midlands (England)	8.3%	10.5%
East of England	9.6%	11.2%
London	15.2%	6.8%
North East (England)	5.1%	4.2%
North West (England)	13.2%	13.8%
South East (England)	15.3%	16.5%
South West (England)	8.3%	10.2%
Wales	5.5%	5.6%
West Midlands (England)	9.8%	10.5%
Yorkshire and The Humber	9.8%	10.7%
TOTAL FLOOR AREA		(SMD=0.056)
Floor area	$87.34m^2$	$84.87m^2$
HABITABLE ROOMS		(SMD=0.188)
less than 4	35.3%	28.4%
4 to 5	45.6%	54.8%
more than 5	19.1%	16.8%
Number of properties	(N=443,574)	(N=147,842)

Note: This table compares a random sample of the population of all properties (column 1) and the sample of repeated property sales employed in the main analysis (column 2). For the continuous covariate TOTAL FLOOR AREA the mean is displayed. For each categorical covariate, percentages of each level are reported. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG. Demographic data are from the ONS.

6.4 Conclusion

In this paper, we examine a specific policy intervention in the UK housing market in 2018 which targeted houses that were the least energy-efficient. This Minimum Energy Efficiency Standard (MEES 2018) came into force in England and Wales on 1 April 2018. It aimed at encouraging landlords and property owners to improve the energy efficiency of their properties which should reduce overall greenhouse gas emission. It restricted the granting and continuation of existing tenancies where the property was not energy-efficient, with an Energy Performance Certificate (EPC) rating of F and G. We examine how transaction prices of properties adjust to the implementation of this policy. We investigate transaction prices of the least energy-efficient properties which are affected by the policy relative to prices of properties that are unaffected.

Our results suggest that prices of properties affected by the MEES 2018 decreased by about £5-9k relative to unaffected ones. The magnitude of this effect compares well to our theoretical priors. Evidence for a decrease of property prices and its magnitude is relevant for at least three reasons: Firstly, market participants do react to climate policies if they have access to relevant information. The size of our estimated effect is in an order of magnitude to be expected considering the fines and exemptions of non-compliance. Hence, we take this result as evidence in favour of semi-strong market efficiency (Fama, 1970), as real estate markets seem to price in the publicly available information about the energy efficiency of the underlying property. Secondly, the existence of this effect suggests that concerns about "transition risk" from climate change (Carney, 2015) are justified as we indeed find adjustments of property prices in response to the implementation of the policy. However, the size of the effect suggests that implications for financial stability were limited. The policy should have pushed only a very small share of existing homeowners that took out mortgage to finance their house purchase into LTV levels where mortgage lenders would incur losses. Finally, the existence of the effect also raises the question of unintended consequefnces for the wealth distribution of homeowners in England and Wales. However, again the magnitude of the estimated effect is relatively small and it seems that the consequences on the wealth inequality of UK homeowners were limited.

There are aspects of climate policy and transition risks that are beyond the scope of this paper. Future research can investigate the real estate market implications of similar policies that are differently calibrated, for example higher thresholds for exemptions, fines or higher thresholds for energy efficiency standards.

A Data

Table 5 provides a description of the variables from the property transaction data set, sourced from the HM Land Registry website. Each row in the data set represents a property sale that took place in England and Wales.[24]

Table 5: Description of the property transactions data set

Variable	Description
ID	Unique transaction reference number (nominal)
PAON	Primary Addressable Object Name. Typically the house number or name (nominal)
SAON	Secondary Addressable Object Name. Where a property has been divided into separate units (e.g. flats), the PAON will identify the building and the SAON will specify the separate unit (nominal)
Street	The street that the property is located upon (nominal)
Locality	Optional locality information, a more specific location of the property (nominal)
Town	The post town of the post address of the property (nominal)
POSTCODE	The postal code of the property (nominal)
Date	Date when the sale was completed, as stated on the transfer deed (date)
Price	Sale price, as stated on the transfer deed (continuous)

Note: This table provides a description of the property transactions data set obtained from HM Land Registry (HM Land Registry, 2014)

Table 6 provides a description of the variables in the data set covering energy performance certificates (EPC) which we used in our analysis. It is sourced from the Ministry of Housing, Communities & Local Government website (Ministry of Housing, Communities & Local Government, 2020b). Each row in the data set represents information regarding the EPC rating that has been issued for a specific building or building unit in England and Wales.

[24]We note that there is zero variation in the category *Park home* in Table 1 for the PROPERTY TYPE variable as there is only one row with that value. In order to resolve the issue of zero variation, this row was deleted.

Table 6: Description of the EPC data set

Variable	Description
LMK KEY	Unique individual lodgement identifier (nominal)
ADDRESS1	First line of the address (nominal)
ADDRESS2	Second line of the address (nominal)
ADDRESS3	Third line of the address (nominal)
POSTCODE	The postal code of the property (nominal)
CURRENT ENERGY RATING	Current energy rating of the property converted into a linear 'A to G' rating scale (ordinal)
CURRENT ENERGY EFFICIENCY	Current energy rating value of the property, ranging from 1 to 100 (discrete)
PROPERTY TYPE	The type of property (Bungalow/Flat/House/Maisonette/Park home) (nominal)
INSPECTION DATE	The date that the inspection was carried out by the energy assessor (date)
LOCAL AUTHORITY	Office for National Statistics (ONS) code, giving the local authority area in which the building is located (nominal)
TOTAL FLOOR AREA	The total useful floor area (continuous)
NUMBER HABITABLE ROOMS	The number of habitable rooms, including any living room, sitting room, dining room, bedroom, study and similar (discrete)
ADDRESS	Field containing the concatenation of ADDRESS1, ADDRESS2, and ADDRESS3 (nominal)
POSTTOWN	The post town of the property (nominal)
CONSTRUCTION AGE	The age band when the building was constructed (before 1900/1900-1929/1930-1949/1950-1966/1967-1975/1976-1982/1983-1990/1991-1995/1996-2002/2003-2006/2007 onwards) (nominal)
TENURE	The tenure type of the property (owner-occupied/rental (private)/rental (social)) (nominal)

Note: This table provides a description of the EPC data set, as obtained from the Ministry of Housing, Communities & Local Government (Ministry of Housing, Communities & Local Government, 2020b).

For completeness, we also provide an overview of the 2011 Area Classification for Output Areas, as described by the Office for National Statistics (2015). We provide a summary of the eight supergroups forming the top tier of the hierarchy. These provide the most generic descriptions of the population in the UK. Descriptions for supergroups that discuss the average are referring to the "average" characteristics for the UK. If not explicitly stated, comparisons (for example higher or lower) are made with the UK as a whole (Office for National Statistics, 2015).

Rural residents The population of this supergroup live in rural areas that are far less densely populated compared with elsewhere in the country. They will tend to live in large detached properties which they own and work in the agriculture, forestry and fishing industries. The level of unemployment in these areas is below the national average. Each household is likely to have multiple motor vehicles, and these will be the preferred method of transport to their places of work. The population tends to be older, married and well educated. An above average proportion of the population in these areas provide unpaid care and an above average number of people live in communal establishments (most likely to be retirement homes). There is less ethnic integration in these areas and households tend to speak English or Welsh as their main language (Office for National Statistics, 2015).

Cosmopolitans The majority of the population in this supergroup live in densely populated urban areas. They are more likely to live in flats and communal establishments, and private renting is more prevalent than nationally. The group has a high ethnic integration, with an above average number of residents from EU accession countries coinciding with a below average proportion of persons stating their country of birth as the UK or Ireland. A result of this is that households are less likely to speak English or Welsh as their main language. The population of the group is characterised by young adults, with a higher proportion of single adults and households without children than nationally. There are also higher proportions of full-time students. Workers are more likely to be employed in the accommodation, information and communication, and financial related industries, and using public transport, or walking or cycling to get to work (Office for National Statistics, 2015).

Ethnicity central The population of this group is predominately located in the denser central areas of London, with other inner urban areas across the UK having smaller concentrations. All non-white ethnic groups have a higher representation than the UK average especially people of mixed ethnicity or who are Black, with an above average number of residents born in other EU countries. Residents are more likely to be young adults with slightly higher rates of divorce or separation than the national average, with a lower proportion of households having no children or non-dependent children. Residents are more likely to live in flats and more likely to rent. A higher proportion of people use public transport to get to work, with lower car ownership, and higher unemployment. Those in employment are more likely to work in the accommodation, information and communication, financial, and administrative related industries (Office for National Statistics, 2015).

Multicultural metropolitans The population of this supergroup is concentrated in larger urban conurbations in the transitional areas between urban centres and suburbia. They are likely to live in terraced housing that is rented – both private and social. The group has a high ethnic mix, but a below average number of UK and Irish born residents. A result of this is that households are less likely to speak English or Welsh as their main language. Residents are likely to be below retirement age. There is likely to be an above average number of families with children who attend school or college, or who are currently too young to do so. The rates of marriage and divorce are broadly comparable with the national average. The level of qualifications is just under the national average with the rates of unemployment being above the national average. Residents who are employed are more likely to work in the transport and administrative related industries. Public transport is the most likely method for individuals to get to and from work, since households are less likely to have multiple motor vehicles available to them (Office for National Statistics, 2015).

Urbanites The population of this group are most likely to be located in urban areas in southern England and in less dense concentrations in large urban areas elsewhere in the UK. They are more likely to live in either flats or terraces, and to privately rent their home. The supergroup has an average ethnic mix, with an above average number of residents from other EU countries. A result of this is households are less likely to speak English or Welsh as their main language. Those in employment are more likely to be working in the information and communication, financial, public administration and education related sectors. Compared with the UK, unemployment is lower (Office for National Statistics, 2015).

Suburbanites The population of this supergroup is most likely to be located on the outskirts of urban areas. They are more likely to own their own home and to live in semi-detached or detached properties. The population tends to be a mixture of those above retirement age and middle-aged parents with school age children. The number of residents who are married or in civil-partnerships is above the national average. Individuals are likely to have higher-level qualifications than the national average, with the levels of unemployment in these areas being below the national average. All non-White ethnic groups have a lower representation when compared with the UK and the proportion of people born in the UK or Ireland is slightly higher. People are more likely to work in the information and communication, financial, public administration, and education sectors, and use private transport to get to work (Office for National Statistics, 2015).

Constrained city dwellers This supergroup has a lower proportion of people aged 5 to 14 and a higher level aged 65 and over than nationally. It is more densely populated than the UK average. People are more likely to be single or divorced. There is a lower representation of all the non-White ethnic groups and of people who were born in other EU countries. There is a lower proportion of households with no children. Households are more likely to live in flats and to live in social rented accommodation, and there is a higher prevalence of overcrowding. There is a higher proportion of people whose day-today activities are limited, and lower qualification levels than nationally. There is a higher level of unemployment in the supergroup. There are no particular industries in which workers are most likely to be employed, but some industries such as information and communication, and the education sector are underrepresented (Office for National Statistics, 2015).

Hard-pressed living The population of this group is most likely to be found in urban surroundings, predominately in northern England and southern Wales. There is less non-White ethnic group representation than elsewhere in the UK, and a higher than average proportion of residents born in the UK and Ireland. Rates of divorce and separation are above the national average. Households are more likely to have non-dependent children and are more likely to live in semi-detached or terraced properties, and to socially rent. There is a smaller proportion of people with higher level qualifications, with rates of unemployment above the national average. Those in employment are more likely to be employed in the mining, manufacturing, energy, wholesale and retail, and transport related industries (Office for National Statistics, 2015).

B Propensity Score Matching

Classification for PSM

For all the implemented methods that are mentioned in this section, we split our data into 70% training and 30% test sets. During the split, we ensure that both sets have the original class ratio of the pre-intervention data, with the EPC rating of at least E being 93.21% of the total size, and the EPC below E of 6.79%. Before applying the classifiers on the training data using as predictors TOTAL FLOOR AREA, PROPERTY TYPE, CONSTRUCTION AGE, TENURE, DEMOGRAPHIC, REGION, and HABITABLE ROOMS, we check whether the model required tuning of any hyperparameters. Logistic regression requires none. However, random forest and boosting have a range of hyperparameter values that can be fine-tuned by implementing 5-fold cross-validation (CV) (Hastie et al., 2009, p. 241-249). Note that the boosting method that we implemented, was a more advanced, modern version called *eXtreme Gradient Boosting (XGBoost)* (Chen and Guestrin, 2016), whose usage of parallel computing makes it extremely faster, than the classical boosting algorithm. The main advantages of XGBoost over the classical boosting method are that it calculates the second partial derivatives of the loss function to get to the minimum of the loss function, and uses advanced L_1 norm and L_2 norm regularization, which has the effect of reducing variance and improving the predictive performance of the fitted model.

As a first step, we define the set of parameter values that we want to evaluate for each model. Then for each hyperparameter set of values, a 5-fold CV is implemented, and the hyperparameter set of values that maximized the CV area under the ROC curve (AUC) score, is selected. One way to interpret the AUC score, is to think that if we randomly choose a positive instance and a negative instance, then AUC represents the probability that the classifier ranks the positive instance higher than the negative instance. For the random forest, the hyperparameter set of values that maximized the CV AUC score was equal to $m = 2$, while for XGBoost was 500 trees, $d = 6$ and $\lambda = 0.01$.

Table 7: AUC values for the classifiers on the test data

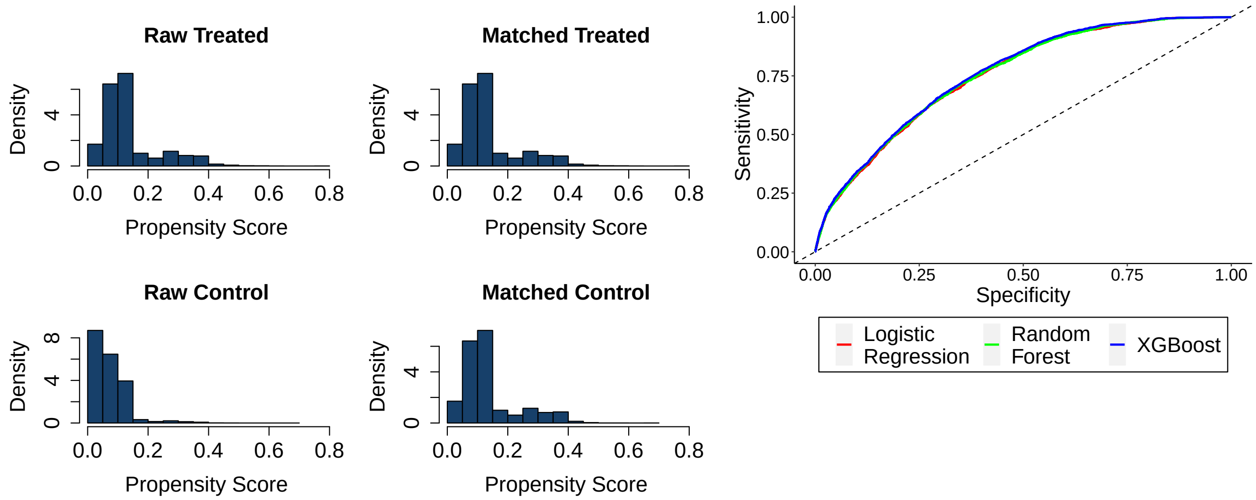
	AUC
Logistic Regression	0.752
Random Forest	0.753
XGBoost	0.759

After fitting a main effects logistic regression model, a random forest with $m = 2$, and XGBoost of 500 trees, $d = 6$ and $\lambda = 0.01$, on the training set, we used the test set to check each model's predictive performance. The value 0.5 was used as a threshold to assign an observation to a class based on the predicted probability. Specifically, if the predicted probability was greater than 0.5, then the property was classified to be class *EPC_LEVEL = Below E*, i.e. belong to the treatment group. On the other hand, if the predicted probability was less than or equal to threshold 0.5, then the property was classified to be

Figure 8: Propensity scores

(a) Histograms of propensity scores before and after matching

(b) ROC curve for classifiers



class *EPC_LEVEL* = *At least E*. One way to check how good each classifier separates the two classes is to plot the ROC curve for different decision cutpoints, as shown in Figure 8a. Sensitivity is defined as the true positive rate, and specificity as the true negative rate, with true positive being a correct positive prediction *EPC_LEVEL* = *Below E* by the classifier, and true negative being a correct negative prediction *EPC_LEVEL* = *At least E*. In situations when classifiers have ROC curves that intersect, one can use the AUC score, which summarizes the overall performance of each classifier. Therefore, from Table 7, we see that XGBoost is slightly superior to the other models with AUC=0.759.

Implementation of Matching Using Propensity Scores

As a next step, we then calculated the predicted probabilities of treatment assignment, i.e. the probability that the property corresponds to *EPC_LEVEL* = *Below E*, in the whole pre-intervention sample of 147,841 rows, using an XGBoost algorithm of 500 trees, $d = 6$ and $\lambda = 0.01$. Then, the *MatchIt* package (Ho et al., 2011) was used to conduct PSM with one-to-one nearest neighbor method, that matches each treated property to only one non-treated property based on their proximity in terms of the propensity scores. More precisely, for $i = 1, 2, ..., n_{\text{treated}}$, where n_{treated} is the number of treated properties in our sample, at the ith matching step. As such, the ith treated property was matched to the closest non-treated property that was not previously matched.

To visualize the balance of the resulting matched data set Figure 8 shows four histograms: the original treatment and control groups, and the matched treatment and control groups. The histograms for the two groups prior to matching on the left side, differ to a great degree, while the histograms after matching on the right side are practically the same. Therefore, graphically, we see that the matching was successful.

Indeed, PSM matching improves the covariate balancing across treatment and control groups. Table 8 compares SMD values of each covariate across our three data sets: in the first column, we show the unmatched data set before we employ matching, and the second and third summarise the matched data sets that come from respectively using the Logistic regression and Boosting models. We note that for both the logistic and boosted methods SMD values are now lower than 0.1 for all variables. We conclude that any difference that existed between the two the treatment and control group prior to matching has been reduced dramatically after implementing PSM. In addition to the slightly higher AUC, it is now clear the boosted method provides superior alignment of covariate balance as measured through SMD. For this reason, we choose the boosted model to continue our analysis confident that the sample is far more comparable than the unmatched data set.

Table 8: Comparing PSM performance with the unmatched data

	Unmatched	Matched (Logistic)	Matched (XGBoost)
	(1)	(2)	(3)
PROPERTY TYPE	0.203	0.028	0.017
CONSTRUCTION AGE	0.924	0.013	0.008
TENURE	0.071	0.048	0.026
DEMOGRAPHIC	0.397	0.037	0.020
REGION	0.181	0.026	0.044
TOTAL FLOOR AREA	0.160	0.013	0.016
HABITABLE ROOMS	0.155	0.040	0.007
Total	2.091	0.205	0.138

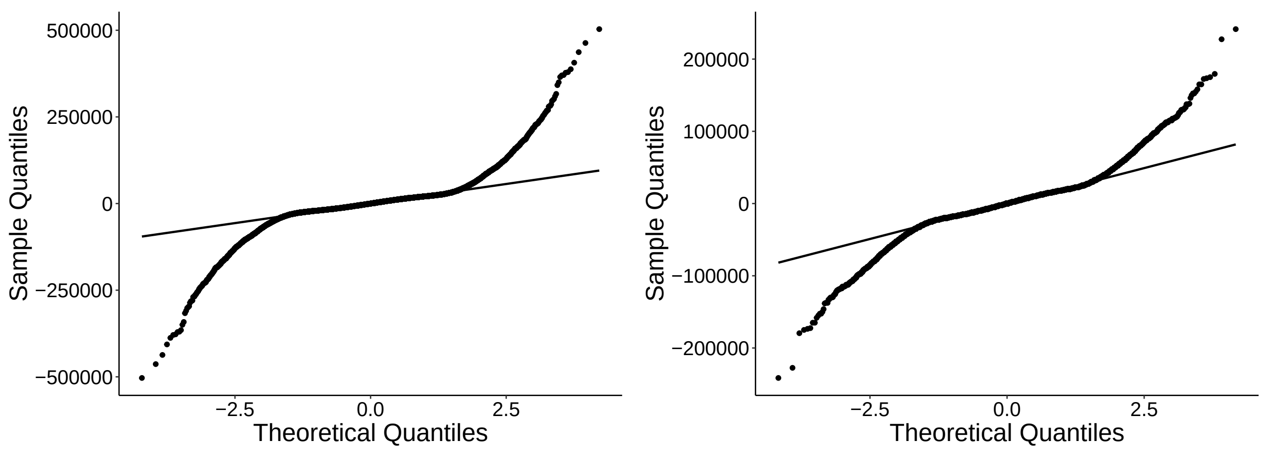
Note: This table shows SMD values for three different data sets. Column 1 shows SMD values in the unmatched sample. Column 2 shows them in the matched sample using a Logistic regression. Column 3 shows them in the matched sample using XGBoost. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG. Demographic data are from the ONS.

C Sensitivity Analysis

On closer inspection of the residuals for the DiD analysis (via equation 7), we note that they have low skewness, but heavier tails than the Gaussian assumption (used in the plm estimation procedure) assumes. This high level of kurtosis (25.3) can be seen in Figure 9. It suggests that the DiD parameters may be unduly influenced by a few outlier properties. To further examine this potential concern, we perform several sensitivity analyses, all of which turn out to verify the robustness of the common trend assumption, the magnitude and significance of our estimated DiD effect, and its direction.

Figure 9: Realised error distribution vs theoretical normal quantiles

(a) Kurtosis 25.3 (b) Kurtosis 9.5



Note: This figure illustrates QQ-plots of the realised error distribution vs theoretical normal quantiles. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

The first sensitivity analysis considers idiosyncratic errors possibly being correlated within properties. This type of correlation can lead to significantly smaller standard errors, narrow confidence intervals, large t-statistics, and small p-values (Cameron and Miller, 2015). Hence, we cluster the standard errors by individual property to allow correlation to exist in the u_{it} errors for the same ith property (Cameron and Miller, 2015, p. 318). We present these adjusted parameter estimates in Table 9. The resulting p-values are only slightly altered, with the effect size exactly the same. Hence, these results indicate that auto-correlation is not a major issue.

Table 9: Effect of the MEES 2018 policy on house prices (clustered S.E.)

Sample	Full Sample	Private Rented
Dependent variable	Price	Price
	(1)	(2)
Post × *Treated*	−9,200	−7,150
	(1,070)	(2,890)
	p < 0.0001	p = 0.0132
Post	51,500	45,000
	(582)	(1,540)
	p < 0.0001	p < 0.0001
Property F.E.	YES	YES
Observations	40,128	5,002
R^2 (*within*)	0.333	0.304

Note: This table shows the results from fitting the regression model in eq. (7) on both the full balanced panel in column (1), and the subsample of properties that are private rented in column (2). Standard errors are clustered at the property level and reported in brackets. All specifications include property fixed effects. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

The second sensitivity analysis considers a higher level of outlier trimming with respect to prices, as we recall that we pre-processed the data by trimming the Price variable at the 1st and 99th percentile of its distribution. To assess the sensitivity to such pre-processing, we now also report the results of an analysis that trims at the 5th and 95th percentile of prices. Under this trimming, the performance of the different PSM algorithms is similar to that of Section 4.4, with the *XGboost* model again performing best. Using this model for PSM and then estimate the GAM, which leads to a deviance explained of 48.7% with a difference-in-smooth p-value of 0.781, again demonstrating little evidence against the common trend assumption. The results from the DiD analysis are presented in Table 10, where correcting for clustered errors gives little difference (similar to the main analysis) and are thus not reported. Interestingly, we see that the R^2 values increase in both the main and rented sub-sample, however, the size of the DiD effect δ decreases slightly from £9,200 to £5,040 (p-value: < 0.0001) and £7,150 to £4,700 (p-value: 0.047) respectively, both statistically significant at the 5 percent level. The increase in model fit can be understood as the analysis now focuses on the centre of the price distribution (see Appendix C Fig. 9). There is still significant but much reduced kurtosis (9.5). Although the magnitude of the estimated effect decreased, the confidence intervals are also reduced and thus the results corroborate the main analysis in suggesting a significant negative impact on prices.

Our last analysis considers a logarithmic transformation of our Price variable in the PSM matched data set. In this case, the GAM model still does not provide significant evidence against the parallel trend assumption in the pre-intervention period. The subsequent DiD analysis demonstrates a negative effect ($\delta_{\log} = -0.0207$, p-value < 0.0001) similar to our main analysis. However, with the logarithmic transformation the

Table 10: Effect of the MEES 2018 policy on house prices (trimmed sample)

Sample	Full Sample	Private Rented
Dependent variable	Price	Price
	(1)	(2)
Post $\times$ *Treated*	$-5{,}040$	$-4{,}700$
	(813)	(2,360)
	$p < 0.0001$	$p = 0.047$
Post	45,000	40,400
	(453)	(1,240)
	$p < 0.0001$	$p < 0.0001$
Property F.E.	YES	YES
Observations	32,864	4,040
R^2 (*within*)	0.449	0.407

Note: This table shows the results from fitting the regression model in eq. (7) using the sample trimmed at the 5th and 95th percentiles of the Price variable. Column (1) shows the balanced panel of all matched properties. Column (2) shows the subsample of properties that are private rented. Standard errors are reported in brackets. All specifications include property fixed effects. Source: House price data are from HM Land Registry. EPC ratings are from the MHCLG.

interpretation of the effect is now relative to the individual house price.

References

Bolton, P. and M. Kacperczyk (2020). Do Investors Care about Carbon Risk? *Working Paper*.

Breiman, L. (2001). Random Forests. *Machine Learning 45*(1), 5–32.

Caliendo, M. and S. Kopeinig (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *IZA Discussion Paper Series*.

Cameron, C. and D. Miller (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources 50*(2), 317–372.

Cannas, M. and B. Arpino (2019). A Comparison of Machine Learning Algorithms and Covariate Balance Measures for Propensity Score Matching and Weighting. *Biometrical Journal 61*(4), 1049–1072.

Card, D. and A. B. Krueger (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review 84*(4), 772–793.

Carney, M. (2015). Breaking the Tragedy of the Horizon-Climate Change and Financial Stability. Speech given at Lloyd's of London.

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

Case, K. E. and R. J. Shiller (1990). Forecasting Prices and Excess Returns in the Housing Market. *Real Estate Economics 18*(3), 253–273.

Chen, T. and C. Guestrin (2016). XGBoost: A Scalable Tree Boosting System. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Crawley, J., P. Biddulph, P. J. Northrop, J. Wingfield, T. Oreszczyn, and C. Elwell (2019). Quantifying the Measurement Error on England and Wales EPC Ratings. *Energies 12*(18).

Croissant, Y. and G. Millo (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software, Articles 27*(2), 1–43.

Department for Business, Energy & Industrial Strategy (2017). The Domestic Private Rented Property Minimum Standard. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/882957/Domestic_Private_Rented_Property_Minimum_Standard_-_Landlord_Guidance_2020.pdf`. Accessed: April 20, 2021.

Department for Business, Energy & Industrial Strategy (2018). Final Stage Impact Assessment: Amending the Private Rented Sector Energy Efficiency Regulations. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/760313/IA_-_Energy_Efficiency__Private_Rented_Property___England.pdf`. Accessed: April 20, 2021.

Dinardo, J. (2010). Natural Experiments and Quasi-Natural Experiments. In S. N. Durlauf and L. E. Blume (Eds.), *Microeconometrics*, pp. 139–153. London: Palgrave Macmillan.

Efron, B. and T. Hastie (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science.* Institute of Mathematical Statistics Monographs. Cambridge University Press.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2016). Heterogeneity in Returns to Wealth and the Measurement of Wealth Inequality. *American Economic Review 106*(5), 651–55.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance 25*(2), 383–417.

Fewster, R. M., S. T. Buckland, G. M. Siriwardena, S. R. Baillie, and J. D. Wilson (2000). Analysis of Population Trends for Farmland Birds Using Generalized Additive Models. *Ecology 81*(7), 1970–1984.

Fougère, D. and N. Jacquemet (2020). Policy Evaluation Using Causal Inference Methods. *IZA Institute of Labor Economics Discussion Paper Series*.

Freyaldenhoven, S., C. Hansen, and J. M. Shapiro (2019). Pre-Event Trends in the Panel Event-Study Design. *American Economic Review 109*(9), 3307–38.

Friedman, J., T. Hastie, and R. Tibshirani (2000). Special Invited Paper. Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics 28*(2), 337–374.

Gale, C. G., A. D. Singleton, A. G. Bates, and P. A. Longley (2016). Creating the 2011 Area Classification for Output Areas (2011 OAC). *Journal of Spatial Information Science 12*(2016), 1–27.

Guin, B. and P. Korhonen (2020). Does energy efficiency predict mortgage performance? *Working Paper*.

Hardy, A. and D. Glew (2019). An analysis of errors in the Energy Performance certificate database. *Energy Policy 129*, 1168–1178.

Hastie, T. and R. Tibshirani (1986). Generalized Additive Models. *Statistical Science 1*(3), 297–310.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer Series in Statistics. New York: Springer.

Herath, S. and G. Maier (2015). Informational Efficiency of the Real Estate Market: A Meta-Analysis. *Journal of Economic Research 20*(3), 117–168.

HM Land Registry (2014). Price Paid Data [Statistical dataset]. `https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads`. Accessed: April 20, 2021.

Ho, D. E., G. King, E. A. Stuart, and K. Imai (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal Of Statistical Software 42*(8), 1–28.

Hong, H., F. W. Li, and J. Xu (2019). Climate Risks and Market Efficiency. *Journal of Econometrics 208*(208), 265–281.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An Introduction to Statistical Learning: With Applications in R*, Volume 103 of *Springer Texts in Statistics*. New York: Springer.

Krueger, P., Z. Sautner, and L. T. Starks (2020). The Importance of Climate Risks for Institutional Investors. *Review of Financial Studies 33*(3), 1067–1111.

Lechner, M. (2010). The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations and Trends in Econometrics 4*(3), 165–224.

Lee, B. K., J. Lessler, and E. A. Stuart (2010). Improving Propensity Score Weighting Using Machine Learning. *Statistics in Medicine 29*(3), 337–356.

Levin, E. and G. Price (2010). Delivering Changes in Housing Wealth Inequality. *Economics paper 6*.

Luellen, J. K., W. R. Shadish, and M. H. Clark (2005). Propensity Scores: An Introduction and Experimental Test. *Evaluation Review 29*(6), 530–558.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business 36*(4), 394.

McCaffrey, D. F., G. Ridgeway, and A. R. Morral (2004). Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies. *Psychological Methods 9*(4), 403–425.

Ministry of Housing, Communities & Local Government (2020a). Energy Performance of Buildings Data: England and Wales [Statistical dataset]. `https://epc.opendatacommunities.org/?/`. Accessed: April 20, 2021.

Ministry of Housing, Communities & Local Government (2020b). Glossary: Domestic EPCs. `https://epc.opendatacommunities.org/docs/guidance#glossary_domestic`. Accessed: April 20, 2021.

Office for National Statistics (2011). Output Area Classification (December 2011) in the United Kingdom [Statistical dataset]. `http://geoportal.statistics.gov.uk/datasets/be648a095a9745998f6961e5ba54e01c_0`. Accessed: April 20, 2021.

Office for National Statistics (2015). Pen Portraits for the 2011 Area Classification for Output Areas. `https://www.ons.gov.uk/file?uri=/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/penportraitsandradialplots/penportraits.pdf`. Accessed: April 20, 2021.

Rambachan, A. and J. Roth (2019). An Honest Approach to Parallel Trends. Working Paper.

Rose, N. L., H. Yang, S. D. Turner, and G. L. Simpson (2012). An Assessment of the Mechanisms for the Transfer of Lead and Mercury from Atmospherically Contaminated Organic Soils to Lake Sediments with Particular Reference to Scotland, UK. *Geochimica et Cosmochimica Acta 82*, 113–135.

Rosenbaum, P. R. and D. B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*(1), 41–55.

Roth, J. (2020). Pre-Test with Caution: Event-Study Estimates after Testing for Parallel Trends. Working Paper.

Rubin, D. B. (2007). The Design versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials. *Statistics in Medicine*, 20–36.

Ryan, A. M., E. Kontopantelis, A. Linden, and J. F. Burgess (2019). Now Trending: Coping with Non-Parallel Trends in Difference-in-Differences Analysis. *Statistical Methods in Medical Research 28*(12), 3697–3711.

Samuelson, P. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review 6*, 41–49.

Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang (2020). Firm-Level Climate Change Exposure. *Working Paper*.

Setoguchi, S., S. Schneeweiss, M. A. Brookhart, R. J. Glynn, and E. F. Cook (2008). Evaluating Uses of Data Mining Techniques in Propensity Score Estimation: A Simulation Study. *Pharmacoepidemiology and Drug Safety 17*(6), 546–555.

Simpson, G. L. (2018). Modelling Palaeoecological Time Series Using Generalised Additive Models. *Frontiers in Ecology and Evolution 6*, 1–21.

Stuart, E. A., B. K. Lee, and F. P. Leacy (2013). Prognostic Score–Based Balance Measures Can Be a Useful Diagnostic for Propensity Score Methods in Comparative Effectiveness Research. *Journal of Clinical Epidemiology 66*(8 SUPPL.8), S84–S90.e1.

UK Statutory Instruments (2015). The Energy Efficiency (Private Rented Property) (England and Wales) Regulations 2015. `https://www.legislation.gov.uk/uksi/2015/962/contents/made`. Accessed: April 20, 2021.

Westreich, D., J. Lessler, and M. J. Funk (2010). Propensity Score Estimation: Neural Networks, Support Vector Machines, Decision Trees (CART), and Meta-Classifiers as Alternatives to Logistic Regression. *Journal of Clinical Epidemiology 63*(8), 826–833.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC Texts in Statistical Science Series. New York: Chapman and Hall/CRC.

Yang, G. and D. L. Moyer (2020). Estimation of Nonlinear Water-Quality Trends in High-Frequency Monitoring Data. *Science of the Total Environment 715*, 136686.