University of Cambridge

Found in the (Random) Forest

April 2021

Abstract

The impact of Airbnb has come under significant scrutiny and this short paper contributes to the literature by looking at Airbnb's effect on house prices in Amsterdam. The key issue is identification due to the likely presence of unobserved confounding factors like tourism demand, which shift housing supply and demand in Amsterdam. We employ Generalised Random Forests to estimate a local average partial effect that comes closest to a causal effect of Airbnb on house prices. These results are compared to the benchmark of a panel data model with time- and area-fixed-effects. The estimated average treatment effects show a nuanced picture of the causal effect of Airbnb presence on local housing demand. Further distance to an Airbnb seems to increase house prices by 0.25% for every 100 meters on average. A 0.019% decrease in house prices per additional listing within 250 meters, on average, seems to suggest a counterintuitive negative effect on local house prices (possibly due to negative externalities). The spillover of Airbnb on neighbouring areas' house prices may be positive, which requires further investigation. The random forset technique also shows that the effect of Airbnb is very heterogenous.

1 Introduction

A brief background on Airbnb
Airbnb is part of the 'sharing economy' and its motivation is to match short-term rental demand with underutilized houses or spare rooms. Because of Airbnb's review systems, potential renters can screen potential landlords. The reduction in trust and quality assurance frictions should ensure a reduction in unused room capacity and so an increase in economic efficiency [15]. This also allows homeowners to rent out their property for short periods of time and so access a direct income stream from their property which should, other things equal, increase their house value.

The platform uses matching technology to reduce the search costs of users and facilitate smooth and safe transactions with its review and reporting mechanism. Airbnb also opens up new areas to tourism with three-quarters of its listings in neighbourhoods typically not covered by the traditional tourist industry. Meanwhile, it offers an authentic and budget rental option for tourists and short-term visitors.

A brief background on Airbnb in Amsterdam
Airbnb was first introduced in Amsterdam in 2008 and has grown rapidly since, with one in 15 dwellings in Amsterdam showing up on an online rental platform such as Airbnb in 2020 [1]. This growth has been accompanied by sharp rises in house prices with house prices increasing 65% in the 5 years to 2018 [14]. This is not necessarily causal, as many other variables could cause the correlation of Airbnb and house price growth.

A concern with Airbnb would be that it benefits non-resident tourists at the expense of residents, which is clearly a public policy problem. This tension between residents and non-residents culminated in 75% of residents in 3 historic regions voting to ban home-rental [17] although this ban was later overturned [1].

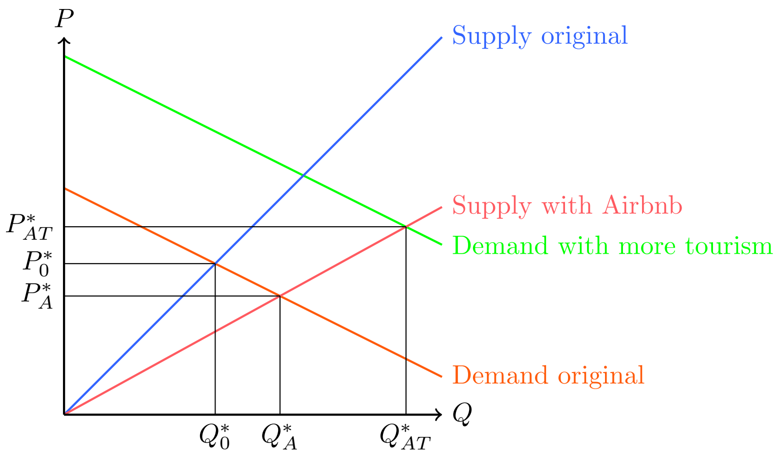
Channels through which Airbnb affects house prices
Prices are fundamentally the result of a supply and demand relationship which is often ignored in hedonic pricing. Yet, this relationship is key to understanding the channels through which Airbnb affects house prices. From this relationship, we know that any variables which shift supply or demand can affect the price. This gives us several clear potential channels as below. These are explored in more detail in section 2.

1. A shift from long-term renting supply to short-term renting causes a reduction in long-term renting stock (rental supply).
2. Airbnb, by reducing frictions of short-term rentals for landlords, can increase the supply of short-term rentals for tourists (rental supply).
3. The increase of profitability of short-term rentals increases demand for houses (housing demand).
4. An increase in the income/wealth of homeowners/other locals in an area due to Airbnb increases their real income and with it demand for housing (housing demand).
5. Negative (non-pecuniary) externalities imposed by Airbnb users on locals reduce the attractiveness of properties for owner-occupiers or long-term renters (rental and housing demand).

It's important to note that while some of these channels such as the externality channel will primarily be a very localised phenomenon, other channels such as the wealth channel will have much more of a dispersed effect across Amsterdam or The Netherlands as landlords do not need to purchase property near their current property. A diversification of risk argument would suggest they may not want to focus their property portfolio in a single area.

Key potential issues
In general, treatment effect identification in a hedonic pricing regression, where the price is modelled as a function of regressors, is challenging. As prices are the result of the equilibrium of supply and demand, any variable that shifts supply or demand also shifts the price. Hence, all variables which would shift the demand or supply equation and which are correlated with the treatment effect must be conditioned on. Any of those variables we don't condition on would be unobserved confounders, which would cause bias in the treatment effect estimator. Estimation of the causal effect of Airbnb on house prices faces all these challenges. The most obvious confounder in Amsterdam's short-time rental market is tourism demand. Short-term rentals are often used by tourists. As the demand for tourism increases, the short-term rental demand increases accordingly (figure 1). The short-term rental rate increases just like the equilibrium quantity of short-term rental. This will happen at the cost of its substitute in supply, long-term rental. The long-term rental rate will increase significantly, while

Figure 1: Short-term rental market with more tourism



the long-term rental equilibrium quantity would decrease even further than it had due to the presence of Airbnb.

The house transaction data set provides us with an advantage for identification, as we can include all necessary time and location dummies without perfectly predicting house prices and Airbnb supply due to within-area-and-time-variation.

Key findings

We initially use panel data methods which suggest that there is a positive effect of distance to the nearest Airbnb on house prices and a negative effect of density of Airbnb's on house prices. It is important to note that in percentage terms these effects are small but in monetary terms the effect is more economically significant.

The random forest we subsequently run supports this analysis and obtains predictions that are close to the area and fixed effect prediction. The mean predicted house price change due to an increase in 100m to the nearest Airbnb is 0.25% while the the mean predicted house price change due to an increase of 10 Airbnbs within 250m is -0.019%. The random forest also finds that the predicted effect of Airbnb density and distance on local house prices varies widely across the set and so to give a single point estimate for the effect of Airbnb density on house prices would be clearly incorrect. Rather, we have a heterogenous treatment effect with the effect depending on the house's specific characteristics.

We also consider the spatial evolution of house prices and the emergence of Airbnb listings across different regions.

A brief literature review

Barron et al (2021) [3] look to answer the same question but for the United States. They use instrumental variables and conclude that on average a 1% increase in Airbnb listings increases rents by 0.018% and importantly house prices by 0.026%. They conclude this is due to the reallocation of the housing stock by landlords and due to the increase in a house's earning potential.

Sheppard and Udell (2016) [16] examine the effect of Airbnb in New York with a hedonic pricing model and find that a doubling of Airbnb listings is associated with a 6-11% increase in property prices. Other estimation methods they consider produce even higher estimates; their difference-in-difference approach estimates the Airbnb treatment (having an Airbnb listing within 300m) increases value by 31%. These results are very strong, especially the difference-in-difference result and should be treated with a high degree of caution.

Horn and Merante (2017) use a fixed-effects model and find that a 1 standard deviation increase in Airbnb listings is associated with an increase in asking rents of 0.4% in Boston [13]. They find that almost half (46%) of Boston's Airbnb listings are listed by operators listing multiple properties. These properties would likely have been long-term rental were it not for the availability of house-sharing services such as Airbnb and supports the hypothesis that Airbnb increases long-term rental prices by enabling landlords to shift from servicing long-term to short-term renters.

Garcia et al (2020) examine the effect of Airbnb in Barcelona, Airbnb's 6th top destination worldwide, at the time of their paper. They run panel fixed-effects and an instrumental variables shift-share approach and conclude that Airbnb increased rents in Barcelona by 1.9% on average. They also find that an increase in Airbnb listings is associated with a decrease in the number of resident households which is consistent with the theory that Airbnb enables landlords to shift from long-term to short-term renters [11].

As explained above, it appears the treatment effect is heterogeneous so to claim a single effect due to the presence of Airbnb would be inappropriate.

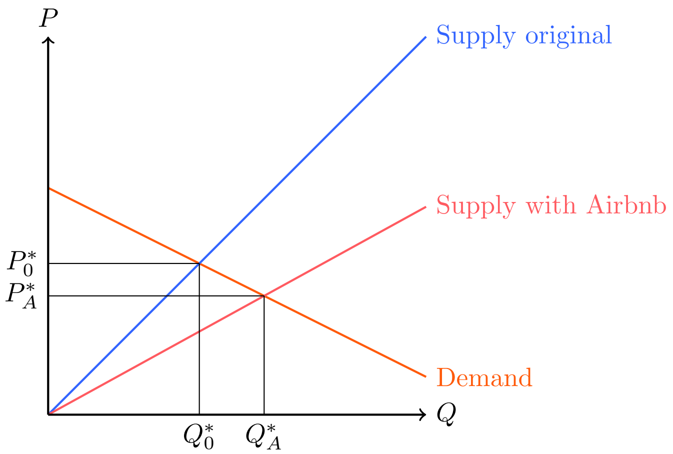
2 Channels

We consider two broad markets: the freehold market and the rental market. The rental market can be further segmented into a short-term rental market mostly for tourists and a long-term rental market for local residents. Hotels, for example, constitute a large part of the short-term rental supply. Short-term and long-term renters have different demands and needs but they both draw from the (same) total housing stock. Traditionally segmentation has existed on the supply side as well as the demand side because of these different needs and because of different legal environments.

The rise of house-sharing is blurring this divide and enables owners of traditionally long-term rental properties to target short-term renters [3]. Airbnb reduces frictions and costs of short-term rental for landlords. These frictions include trust and quality assurance that traditionally would have made short term house rental very difficult [9]. This friction reduction is likely to increase the marginal propensity of homeowners to reallocate housing from the long-term to the short-term rental.

This cost reduction is reflected in an increased supply of short-term rentals at any rent in a perfectly competitive market. In figure 2 we demonstrate this effect by flattening the supply curve for short-term rentals in the quantity-price-space. The equilibrium price of short-term rentals decreases, while the quantity of short-term rentals increases.

Figure 2: short-term rental market



Because of the inelastic nature of the housing stock, this increase in short-term rentals can come from a reduction of long-term rentals, or a reduction in non-rented property. The short-term supply of rentals is nearly inelastic so long-term rental supply has to decrease by approximately the increase in short-term rental supply. Long-term rental supply decreases at every price as the cost of short-term rental decreases with the introduction of Airbnb as demonstrated in figure 3. Local residents in long-term rentals will therefore need to pay a higher equilibrium rental price, while the total quantity of long-term rentals is reduced. The marginal propensity of homeowners to reallocate housing from the long- to the short-term rental market will determine the quantity of replaced housing [3].

The consequences of these changes for the housing market (for purchases) are obvious.

The reduction in short-term renting costs increases the value of potential renting opportunities which increases housing demand. The increase in wealth and income of landlords translates into a greater ability to finance further property investment and so demand and so property prices.

Figure 3: Long-term rental market

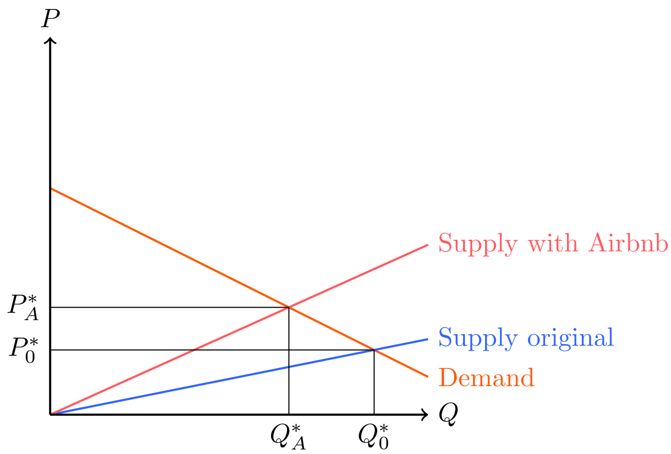
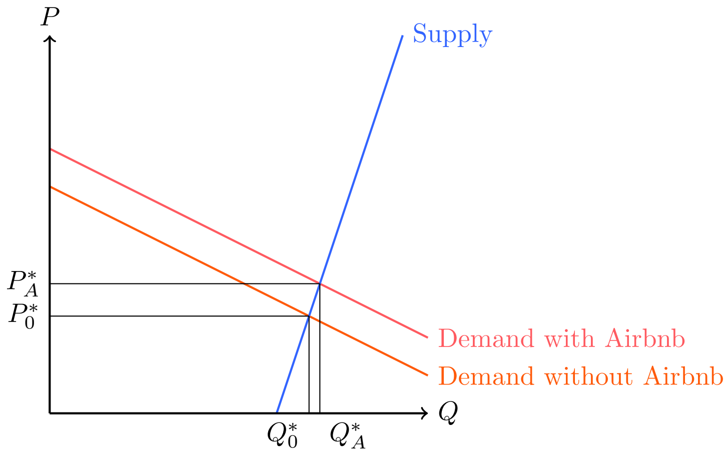


Figure 4: Housing market (for purchase)



The increase in tourists ushered in by Airbnb increases local economic activity and so local incomes and housing demand. These tourists require amenities and services which require land, increasing land demand and so property prices.

Figure 4 demonstrates the increase in demand for housing due to Airbnb. As local housing supply is quite inelastic, the presence of Airbnb should result in an increase in house prices according to the channels covered so far.

For this argument to not go through we would have to have strong negative externalities as a result of local Airbnb's. These externalities could include noise and congestion, can be very large and can outweigh Airbnb's other positive price effects, at least locally. As these externalities are highly local, Filippas and Horton argue that while individual owners will oversupply the market, if the decision is left instead to building owners they will internalise the externality and supply the efficient amount of house sharing [10].

Outside the positive and negative externalities that Airbnb has on house prices, there may be an interaction between Airbnb and the market microstructure. Genesove and Mayer (1997) [12] demonstrate that seller motivation (via higher LTV) leads to higher asking prices. This results in a longer time-on-market (TOM) but conversely the seller achieves a higher sales price. The authors estimate an annualised return of 20 percent for those sellers setting a higher asking pricing. However, Dubé and Legros (2016) [8] have shown that, after controlling for endogeneity, there is in fact a negative relationship between TOM and the sales price reflecting negative information about the property that is not otherwise captured by the housing characteristics in hedonic pricing model. Airbnb may lead to changes in the sales dynamic. By generating income from short-term lets on Airbnb, the seller can reduce the cost of an extended TOM. In this fashion, Airbnb may shift bargaining power to the seller and result in higher sales prices. However, identification remains challenging as higher TOM may motivate sellers to avail of Airbnb to mitigate the costs associated with TOM.

3 Data

The analysis was undertaken using microdata on housing transaction covering the period from 2000-2018. This was supplemented by data from http://insideAirbnb.com/ which we used in the spatial analysis.

The dataset contains information on sales price, distance to nearest Airbnb listing and the number of Airbnb listings within a 250 metres radius along with housing characteristics to use as control variates. There is known measurement error within the Airbnb variables due to the fact that listing locations are only accurate within 100 metres.

Table 1 contains descriptive statistics for the key independent and dependent variables. We identified a number of outliers, where for example the room size was ≥ 20 or 0, and these were removed.

Table 1: Descriptive Statistics

Variable	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Outcome Variables							
price	108,441	301,044	228,108	50,000	173,937	340,000	2,500,000
logprice	108,441	12.447	0.532	10.820	12.066	12.737	14.732
Treatment Variables							
distance (since 2008)	68,773	433.17	1,076.93	0.06	19.71	207.10	8,842.94
density (since 2008)	68,773	69.38	105.688	0	1	98	685
density (since 2000)	108,441	44.00	90.55	0	0	37	685
Case 2 Variables							
asking_price	108,441	330,851	4,308,635	25,000	179,000	349,000	999,999,999
logask	108,441	12.480	0.533	10.127	12.095	12.763	20.723
tom	108,441	117.164	183.045	0	22	134	3,822
Covariates							
rd_x	108,441	120,923.90	3,157.52	112,30	118,92	122,99	132,34
rd_y	108,441	486,087.90	2,494.16	477,02	484,88	487,61	493,07
construction_period	108,441	4.195	2.694	1	2	7	9
garden	108,441	0.273	0.446	0	0	1	1
size	108,441	86.667	42.957	25	60	100	1,185
volume	108,441	246.021	142.040	55	162	284	4,740
rooms	108,441	3.248	1.387	0	2	4	103
wtype	108,441	−0.618	1.071	−1	−1	−1	5
parking	108,441	0.104	0.305	0	0	0	1
monumentalstatus	108,441	0.031	0.174	0	0	0	1
buyerpaysorfree	108,441	1.036	0.187	1	1	1	2
quality	108,441	14.395	1.766	2	14	14	18

Figure 5 shows the correlation between our variables. The key correlations to notice is that there a weak positive correlation between the density of Airbnb's and log price while there is a weak negative correlation between the distance to the nearest Airbnb and log price.

Figure 6 shows the evolution of house prices using a spatial heat map. To aid visualisation, we have grouped the data into blocks of years. The figure shows the concentration of houses prices by neighbourhood. We have then overlaid the average number of Airbnb listings by neighbourhood, this data is available from 2008 to 2018. The Airbnb data highlights the increasing density of listings over time and the heterogeneity across local areas and presents the possibility to determine local treatment effects.

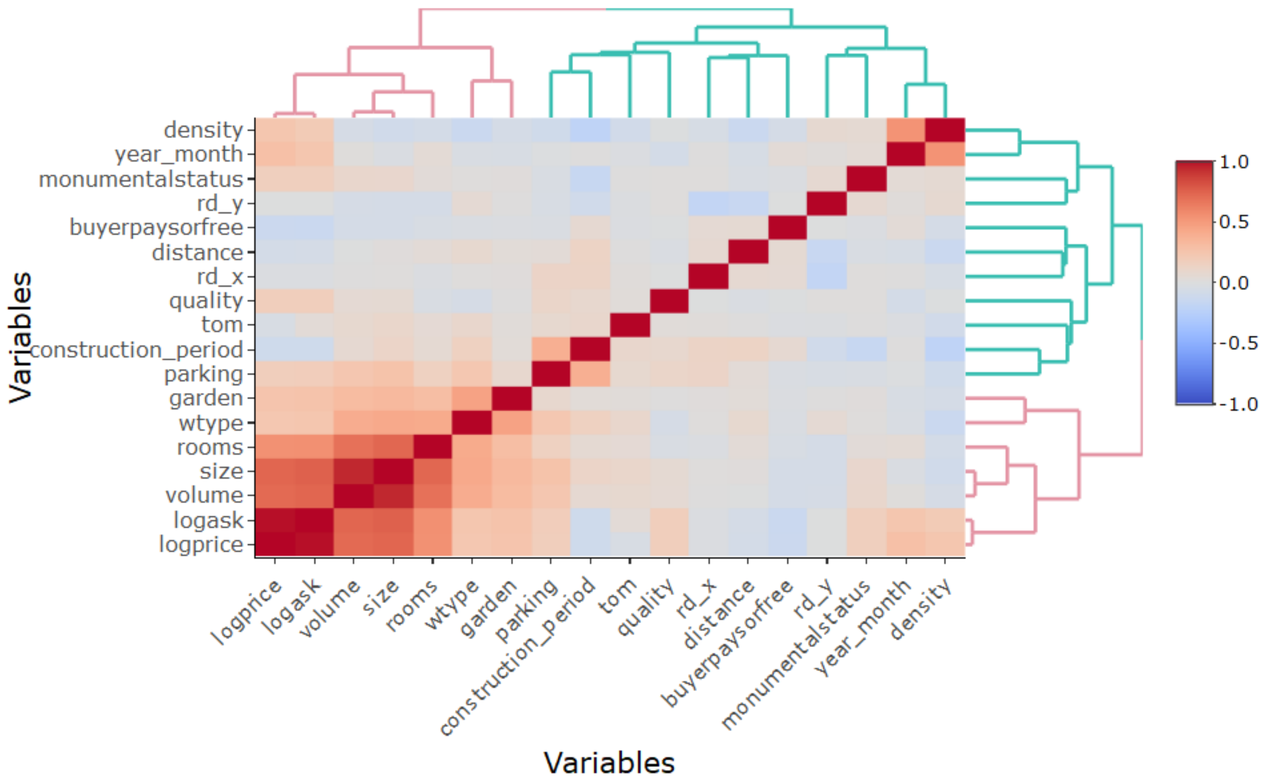


Figure 5: Correlation between variables

The heatmap above shows a weak positive correlation between density and log price; a weak negative correlation betweeen distance and log price. It also shows varying patterns of correlation between covariates.

4 Methodology

Our preferred methodology is Generalised Random Forests, following [2], a nonparametric statistical estimation technique to determine local treatment effects. In this section, we explain the identifying assumptions and estimation technique. We compare the generalised random forest to a benchmark panel model.

4.1 Identification

The price of local houses is determined in a supply and demand equilibrium. As the presence of Airbnb changes, both demand and supply can change with it through one or multiple of the channels we discussed. Our causal effect of interest is the new equilibrium price after the introduction of Airbnb in the local housing market. We measure Airbnb presence with house distance to the closest Airbnb or density within a 250 meter radius. In order to identify the causal effect of Airbnb, we postulate a conditional unconfoundedness assumption 1. Write P_{it} for house price, A_{it} for Airbnb presence (including density and distance) and X_{it} for all covariates as explained in section 3.

Assumption 1 (Conditional Unconfoundedness) $P_i \perp\!\!\!\perp A_i = a | X_i$ *for all* $a \in \mathcal{A}$

In other words, the unconfoundedness (or conditional independence) assumption 1 means that conditional on all observed covariates X_i, the treatment A_i is exogenous. This condition imposes some strong requirements. There may be no unobserved variables that effect both houses price and Airbnb presence. Nonetheless, we have good reason to believe that this assumption is satisfied in our data set. In general, X_i may include all observed confounders, including spatial coordinates and time. We include all of these variables and allow any interactions as soon as we model house prices and Airbnb presence nonparametrically. The time variable will capture any unobservables that affect each local demand and supply for houses equally. Such variables include real income, the interest rate, the location's demographics, etc. Even if unobserved variables interact with local conditions, like the time- and location-specific demand for tourism, our approach works. We nonparametrically model any such interactions automatically. For this reason, there would be no benefit in an inclusion of interest rate or tourism demand data (as it is available for example via CBS Statline).
Our benchmark panel model does not account for all these linearities. At best it linearly accounts for covariates with time and 6-digit-post-code fixed effects.

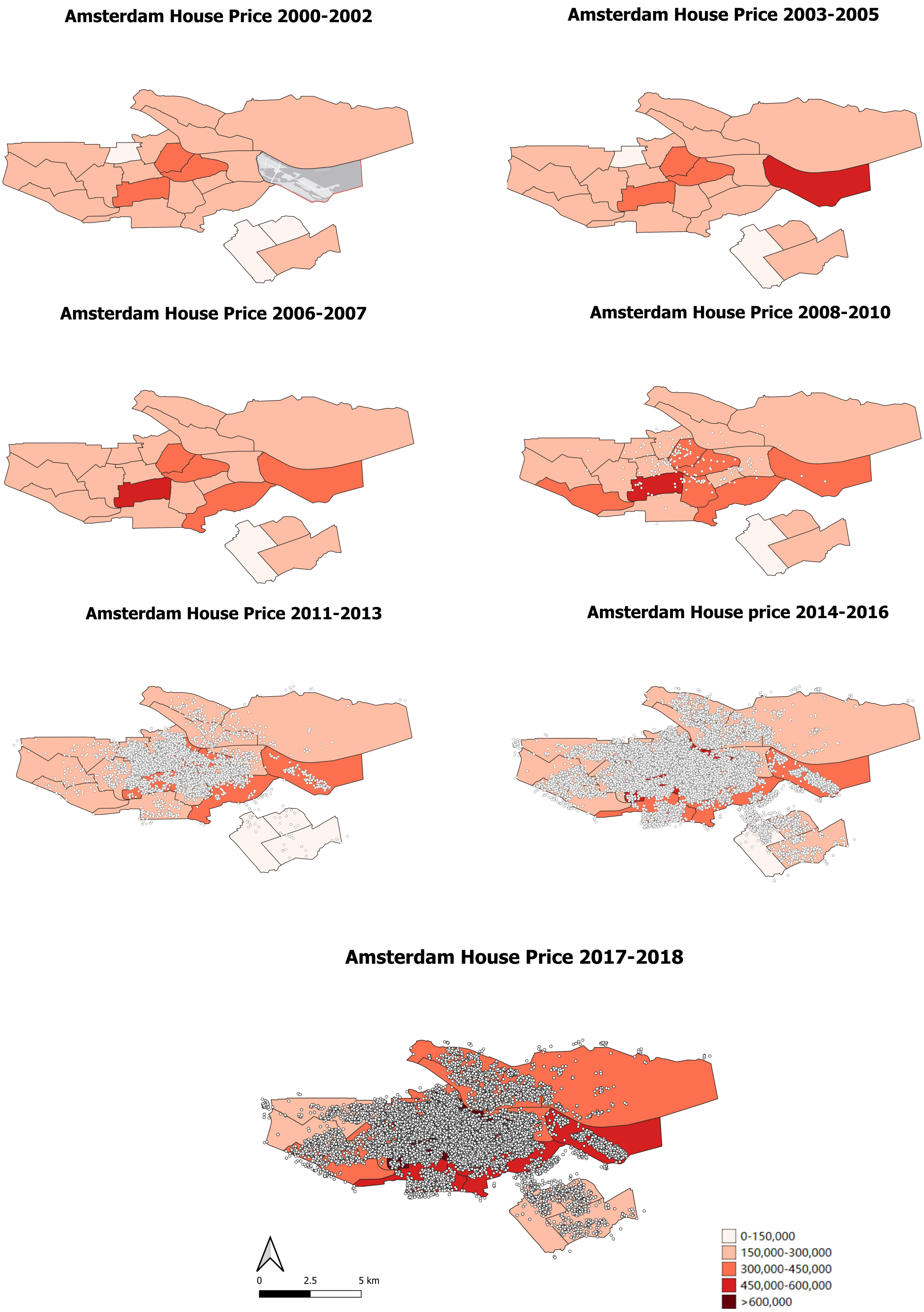


Figure 6: House Price Heat Maps

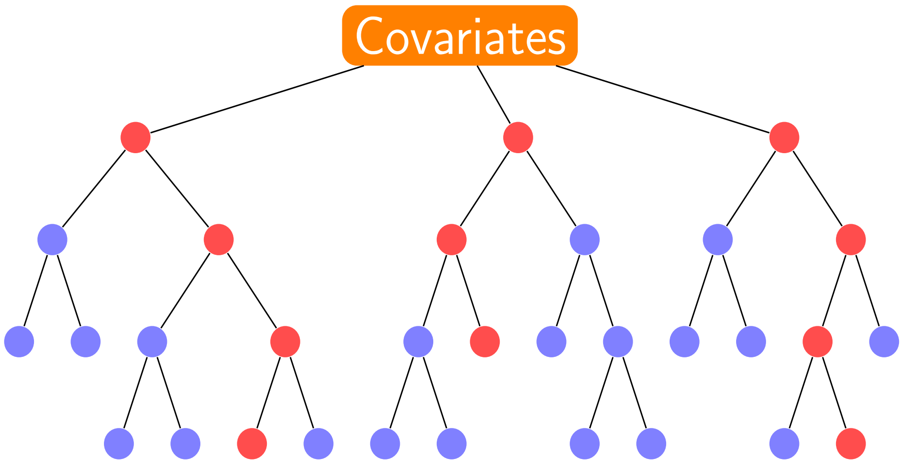


Figure 7: Illustration of a decision tree

4.2 Local estimating equation

In our opinion, the local average partial effect of Airbnb presence on housing prices, $\theta(X_i)$, comes closest to a causal effect of Airbnb on house prices. The local average partial effect has the advantage that no parametric modelling is required. Our local estimating equation is defined as $\psi_{\theta(X),\nu(X)}$. It contains the local treatment effect $\theta(X_i)$ as well as a nonparametric nuisance $\nu(.)$. It is a function of the observed data, where $O_i = \{P_i, A_i, X_i\}$.

$$\mathbb{E}[\psi_{\theta(X),\nu(X)}(O_i)|X_i = x] = 0 \tag{1}$$

$$\psi_{\theta(X),\nu(X)}(O_i) = \theta(X_i) - \frac{Cov(P_i, A_i|X_i = x)}{Var(A_i|X_i = x)} \tag{2}$$

Estimation of local moments (instead of global moments) helps us with identification. We estimate the model fully nonparametrically, but retain the asymptotic normality of the estimated treatment effects. For this reason, we are able to provide confidence intervals of our estimates.

4.3 Random Forest Algorithm

The generalised random forest algorithm for local moment estimation consists of three main steps. The forest consists of a number of decision trees. We depict a decision tree in figure 7 Each decision tree iteratively chooses the best possible splits over the set of covariates to predict an outcome. Decision trees tend to have low bias, but high variance in their predictions [7].
The random forest averages over many decision trees to obtain a low-bias estimator, which also has lower variance [4]. We estimate the price and Airbnb density and distance from all covariates with a random forest of 500 trees and a minimum leaf size of 10 (at least 10 observations in each final node of each tree). By allowing the random forest algorithm to select any covariates and their interactions which matter (in the sense that they predict house price or Airbnb presence), we do not need to include global confounders of house prices, which are the same for all houses at one point in time. In a third step, we estimate a forest to estimate the treatment effect by locally regressing the residual house price on residual Airbnb presence. The algorithm will find similar observations (in terms of covariates) with different Airbnb presence. From these similar observations, we estimate the local moments and hence the local treatment effect with its standard error. As we estimate both the outcome and the treatment conditional on covariates and obtain their residuals, we will ultimately get doubly robust estimates [6]. Intuitively, we find all variables that affect either house price or Airbnb, so we are twice as certain that we find all confounders in the set of covariates. Doubly robust estimates often have better small-sample bias properties and behave asymptotically normally.

4.4 Benchmark Panel Model

Instead of this flexible approach for local average partial effect estimation, we could have used a simple panel model with location (4- or 6-digit post-code) fixed effects α_i, time fixed effects ζ_t and linear covariates to obtain an average treatment effect θ.

$$\mathbb{E}[\psi_{\theta(X),\nu(X)}(O_{it})|X_{it}, A_{it}] = 0 \tag{3}$$

$$\psi_{\theta(X),\nu(X)}(O_{it}) = \theta - \frac{Cov(Y_{it}, A_{it}|X_{it}, \alpha_i, \zeta_t)}{Var(A_{it}|X_{it}, \alpha_i, \zeta_t)} \tag{4}$$

A more common way to write this type of model is in form of a linear equation. This simpler identification strategy is our benchmark.

$$Y_{it} = \alpha_i + \zeta_t + \theta A_{it} + \beta^T X_{it} + U_{it}, \quad \mathbb{E}[U_{it}|\alpha_i, \zeta_t, A_{it}, X_{it}] = 0 \tag{5}$$

$$\tag{6}$$

We estimate the benchmark panel model with the treatments Airbnb density and distance from 2008 onwards, as distance is available only since 2008. We estimate the generalised random forest with the treatment distance from 2008 onwards. We also estimate the panel model with treatment Airbnb density using the data going back to 2000 to see if the pre-Airbnb time periods can provide a different perspective on the causal effect of Airbnb (while we continue to condition on time and 4-digit or 6-digit post-code dummies). A random forest with treatment Airbnb density is run on the same data going back to 2000.

5 Analysis

Our results point to a nuanced causal effect of Airbnb on house prices. If we increase distance from the nearest Airbnb by 100 meters, the population mean predicted change in house prices is 0.25% with our preferred method, the generalised random forest. As there are 10 more listings within a 250 meter radius, the house price is predicted to decrease by 0.19% on average. These results are statistically and economically significant.

5.1 Panel Model Results

We always condition estimation on observed covariates, which capture house characteristics, as described in the Data section 3. The heatmap in figure 5 showed various patterns of correlations of the covariates with each other and log price. Hence, their inclusion will account for some of their possible confounding for the causal effect of Airbnb on house price. In the panel model, we included the covariates linearly.

For the OLS model the Airbnb distance and density variables improve the R^2 by 3.89% from 70.03% to 73.92%. For the area and time fixed effect model, however, Airbnb treatment variables only improve the R^2 by 0.18% from 82.75% to 82.93%. Most variation in house prices is explained by house characteristics and 6-digit-post-code- and time-fixed-effects (82.75%). This leads to the tentative conclusion that Airbnb's presence is not a major driver of Amsterdam's house prices, despite the contrary public narrative.

It is worth noting that we might have overstated the irrelevance of Airbnb if we included many area- and time-fixed-effects unnecessarily, so we will compare the size of these results to those from the generalised random forest, which chooses only covariates that predict either house price or Airbnb density/distance.

In table 2, the estimated average partial effect varies significantly across specifications. All estimated coefficients are strongly statistically significant, which typical in large microeconomic datasets with rich variation in regressors. The difference between the estimates does not represent a statistical margin of error but stems from difference in the models. In the pooled OLS specification, it appears that house prices decrease by 0.33% for each additional 100 meters of distance from the nearest Airbnb. For each additional 10 listings in a 250 meter radius we predict an increase in housing prices of 0.99%. The estimated effect of density does not change much between the pooled OLS and post-code-fixed-effects models, as it is 1.05% in the 4-digit post-code model and 0.91% in the 6-digit post-code model. The estimated effect of distance however is much reduced in absolute value, as it is only 0.04% in the 4-digit post-code model and 0.01% in the 6-digit post-code model. The time fixed-effects model also has a reduced estimate of the distance in absolute size at -0.16%. Notably, the density effect estimate is much lower at 0.26%. This reduction is expected as we expected tourism demand and other macro, time-specific effects like the interest rate, to account for much of the positive correlation between Airbnb density and house prices. We also notice a further reduction of the density estimate to 0.19% when we use both 6-digit post-code- and time-fixed-effects. While the fixed-effects did not lead to a bias reduction by themselves in model (2) and (3), it appears that the combination of post-code and time information allows us to reduce bias even further than what the sum of their individual bias reductions may suggest.

The most noteworthy effect in this table is the change in sign of the distance effect in model (5). Now we estimate that further distance from the Airbnb by 100 meters increases house prices by 0.17%. We interpret this result as the consequence of negative externalities on local housing prices. Airbnb inevitably facilitates short-terms rental for landlords, which improves its supply and its equilibrium quantity. Nonetheless, the negative externalities of Airbnbs on owner-occupiers and long-term renters are so large that the local demand of housing decreases. This is an important result, which we will seek to confirm with the generalised random forest. While the local house price drops, long-term rental will still be displaced by short-term rental. Long-term renters would look for accommodation further away from Airbnbs, which will lead to demand increases in other parts of the city which still have less Airbnbs. In an equilibrium model of all houses demand and supply in Amsterdam, the secondary local demand displacement effect would apply everywhere (to different degrees). Hence, the overall price of houses may still increase as a result of Airbnb presence, even if its local effect on house prices is negative. The investigation of these secondary displacement effect on demand is beyond the scope of this paper and remains a promising next step to understand the dynamic micro effect of Airbnb presence on macro house prices (in the entirety of Amsterdam).

Table 2: Average treatment effect estimate $\hat{\theta}$ for Airbnb distance and density with the panel model and data since 2008

	Dependent variable:				
	logprice				
	Pooled OLS	pc4-FE	pc6-FE	Time-FE	pc6- & Time-FE
	(1)	(2)	(3)	(4)	(5)
distance (in 100m)	-0.0033^{***}	-0.0004^{***}	0.0001^{***}	-0.0016^{***}	0.0017^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
density (in 10 listings)	0.0099^{***}	0.0105^{***}	0.0091^{***}	0.0026^{***}	0.0019^{***}
	(0.00012)	(0.00012)	(0.00012)	(0.00009)	(0.00011)
Observations			68,770		
R^2	0.739	0.750	0.796	0.782	0.829
Res. SE (df=68745)	0.272	0.272	0.275	0.246	0.260
F Stat (df=24; 68745)	$4,341^{***}$	$4,341^{***}$	$8,171^{***}$	$1,286^{***}$	$8,823^{***}$
Note:					*p<0.1; **p<0.05; ***p<0.01

R-squared increases from 73.9% to 82.9% from Pooled OLS to area and time fixed effects, the effect of distance rises from -0.33% to 0.17% and the effect of density falls from 0.99% to 0.19%.

The results in table 3, for which we used data since 2000, confirm the results in table 2. With pooled OLS, the predicted effect of density is large (1.44%) and it remains so with post-code fixed-effects in model (2) (1.46%) and model (3) (1.24%). Once we condition on time effects, the estimated effect is drastically reduced to 0.38%. When we then condition on both location and time fixed-effects we find the estimated effect is reduced further to 0.18%. These results indicate that without using these fixed-effects we will be mistakenly attributing price changes to Airbnb.

5.2 Generalised Random Forest Results

The manual inclusion of many fixed-effects may be an additional reason for the low explanatory power of Airbnb presence. In the generalised random forest we overcome this issue by letting our algorithm choose all nonlinearities and interactions that predict house prices and Airbnb presence. We however avoid the inclusion of unnecessary fixed-effects. For example, if January 2012 and February had the same time fixed effect, the random forest would automatically include only one fixed effect for two periods.

To address these incoherent results, we turn to random forests. Via k-fold cross validation we chose the optimal minimum node size for prediction of house prices and Airbnb density and distance. We chose 10 as the minimum node size and used over 500 trees in the forest with k-fold cross validation, which is a method for an optimal selection of complexity in machine learning and nonparametric statistics [5]. With the covariates listed

Table 3: Average treatment effect estimate $\hat{\theta}$ for Airbnb density with the panel model since 2000

	Dependent variable:				
	lpdm				
	Pooled OLS	pc4-FE	pc6-FE	Time-FE	pc6-FE & Time-FE
	(1)	(2)	(3)	(4)	(5)
density (in 10 listings)	0.0144***	0.0146***	0.0124***	0.0038***	0.0018***
	(0.00012)	(0.00013)	(0.00011)	(0.00008)	(0.00008)
Observations	108,437				
R^2	0.714	0.733	0.790	0.777	0.832
Res. SE (df=108413)	0.284	0.275	0.244	0.251	0.218
F Stat (df=23; 108413)	11,790***	6,663***	2,051***	13,375***	1,362***
Note:					*p<0.1; **p<0.05; ***p<0.01

test

in the Data section (including coordinates rd_x and rd_y instead of post-code as random forests allows for this additional flexibility), we can predict 88% of variation in log prices, 92.4% of variation in Airbnb density, and 98% of variation in Airbnb density. While such good performance is encouraging, it tells us nothing about the estimated causal effects. The estimated local average partial effects are summarised below.

Just like in table 2, the estimated density average partial effect appears to be mostly negative with a mean of 0.019% decrease in houses prices for an additional listing, while for each additional 100 meter distance from an Airbnb the house price appears to increase by 0.25%. Our generalised random forest estimates in table 4 appear much more realistic compared to the unrealistically large estimates of the panel model. This table also suggests that the effect of Airbnb on house prices is heterogenous with the predicted effect of the distance variable ranging from -4.6% to 5.1% and the predicted effect of the density variable ranging from -2.8% to 6.4%.

Table 4: Predicted local effect of density and distance on log home price in Amsterdam

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
distance (in 100m)	68,771	0.0025	0.0087	−0.0460	−0.0027	0.0068	0.0514
density (in 10 listings)	108,437	−0.0019	0.0056	−0.0281	−0.0057	0.0014	0.0641

The table above shows that the an extra 100m distance from an Airbnb increases predicted prices by 0.25% on average with a large interquartile range of 0.95%. An extra 10 listings within 250m decreases predicted prices by 0.19% with an interquartile range of 0.71%. This range is clearly large and important.

A useful form to illustrate the estimated treatment effects in our sample is an empirical cumulative distribution plot as in figures 8 and 9. The heterogeneity of the estimated effect shows in these illustrations, where estimated effects are on the x-axis. For example, we see that approximately 40% of the density effects are negative at the 10% level of significance, as the estimated confidence interval of the treatment effect with its 5% and 95% percentile lies to the left of 0.

As discussed in the methodology, an attractive feature of the Generalised Random Forests approach ability to capture heterogenous treatment effects across the different housing characteristics. To illustrate this point, we show a spatial plot of the average treatment effect for the two Airbnb variables by region, see Figure 11.

In Figure 12 we plot the treatment effect associated with the distance to the nearest Airbnb by region and the mean sales price in the region. From this, we can identify a clear linear relationship between the local treatment effect and the mean sales price. Again, the benefits of a Generalised Random Forests estimation technique can successfully cater for these heterogenous effects.

Comparing the time and fixed-effects regression and the random forest method we can see that the point esti-

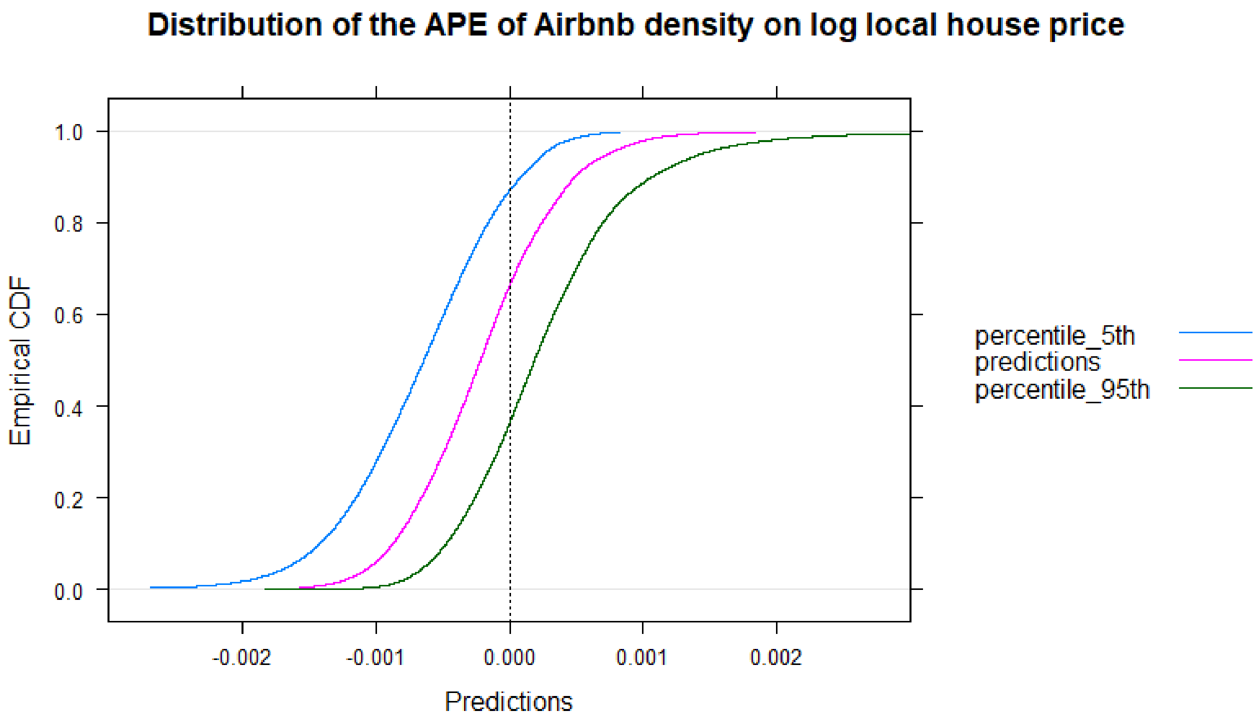


Figure 8: Random forest treatment effect estimate for density

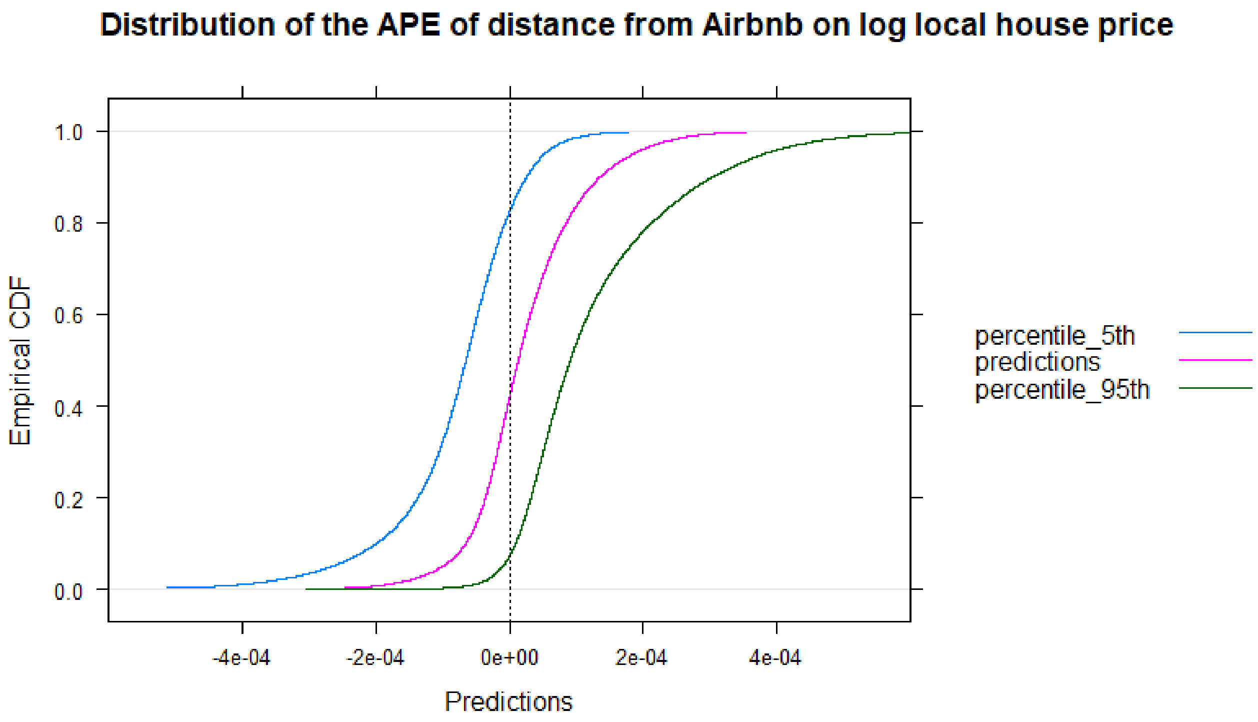


Figure 9: Random Forest treatment effect estimate for distance

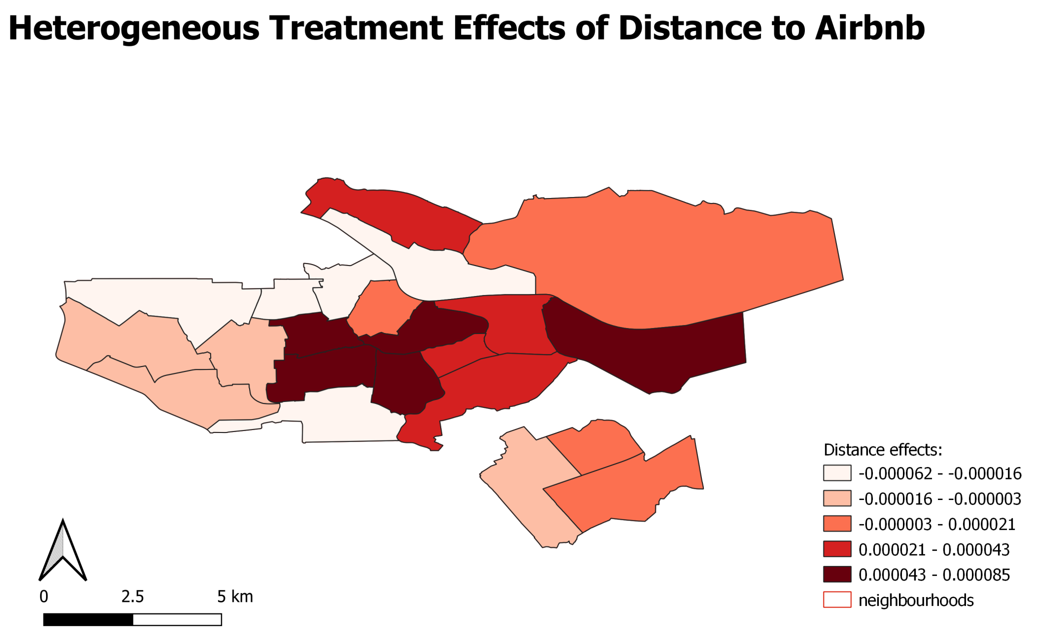


Figure 10: Heterogenous Treatment Effects - Airbnb Distance

The figure above shows that there are several regions with relatively high positive distance effects and several with relatively high negative distance effects.

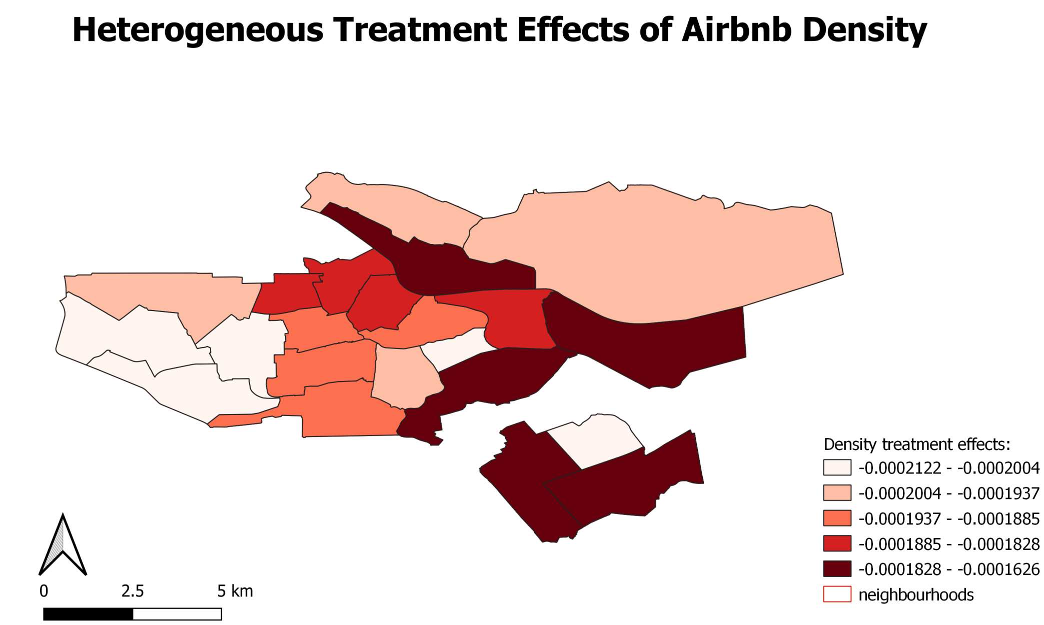
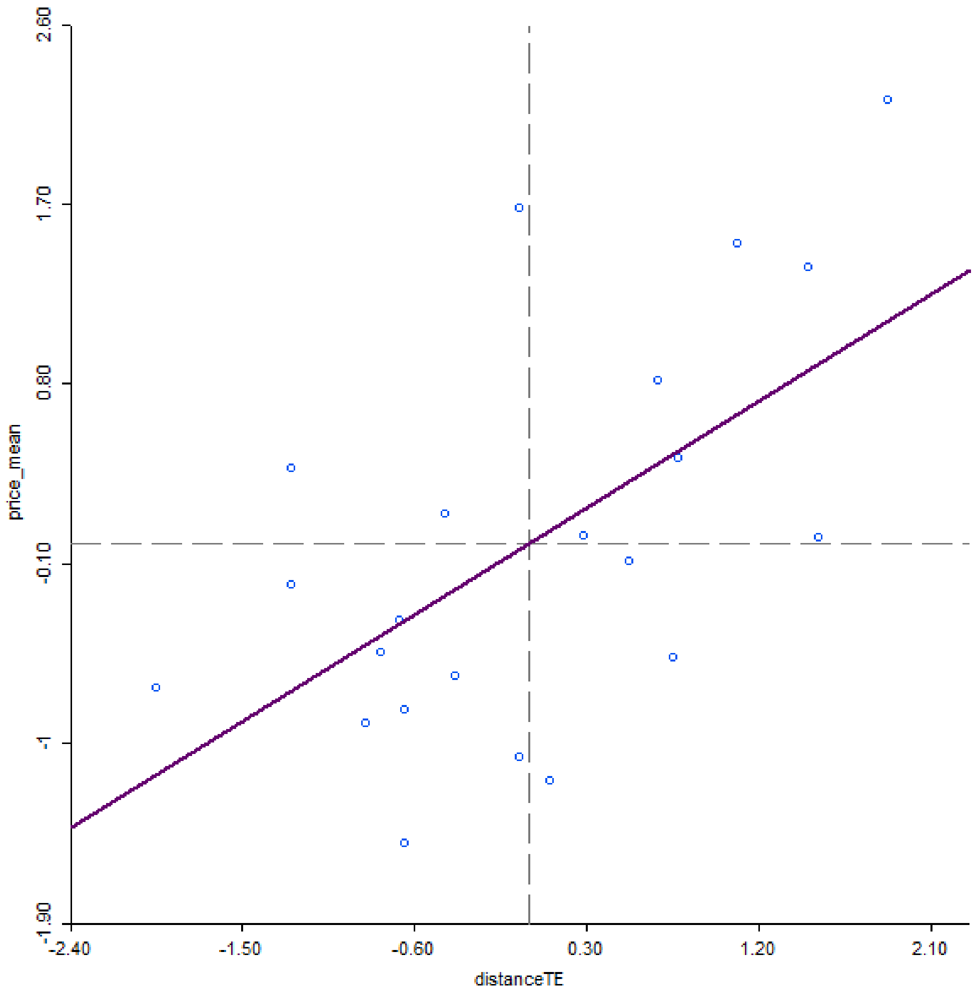


Figure 11: Heterogenous Treatment Effects - Airbnb Density

The figure above shows that the density treatment effect is negative for all neighbourhoods.



#obs	R^2	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
22	0.352	0	0.176	0	1	0.594	0.180	3.300	0.004
0	0	0	0	0	0	0	0	0	0
22	0.352	0	0.176	0	1	0.594	0.180	3.300	0.004

Chow test for sel/unsel regression subsets: can't compute

Figure 12: Treatment Effects - Airbnb Distance by region

The figure above plots the treatment effect for each region against the mean sales price by region, and show a strongly positive relationship between the two.

mates for the effect of distance are of the same sign (positive) for both. The point estimates for the effect of density are of the same small magnitude so while they are of different signs, the difference is small.

The time and fixed-effects regression improves drastically on the pooled OLS but it does not fully obtain the random forest results.

5.3 Effect of Airbnb on house price and time-on-market equilibrium

We investigated the causal effect of Airbnb presence on the time-on-market equilibrium using Generalised Random Forests technique. As with earlier analysis the effect is heterogeneous and for the most part the effect is not statistically significant. To illustrate the lack of statistical significance, we have provided empirical CDF plots, see Figure 14. These plots show that almost none of the predicted effects of density or distance on TOM are statistically different from zero using 2-sided tests at the 10% level.

While we hypothesised that the presence of Airbnb could increase rental options while someone is looking to sell and so increase their ability to sustain a non-sale for an extended period of time leading to a greater TOM, this hypothesis is not supported by this evidence.

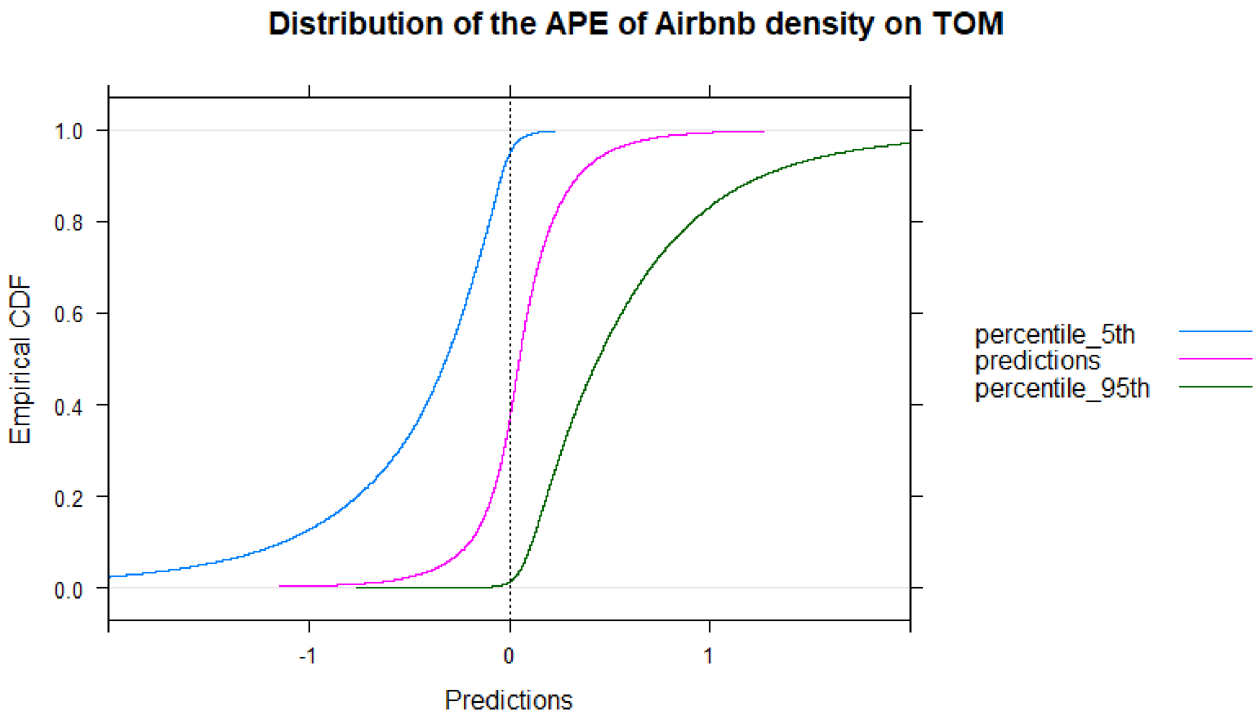


Figure 13: Random forest treatment effect estimate for density on time-on-market

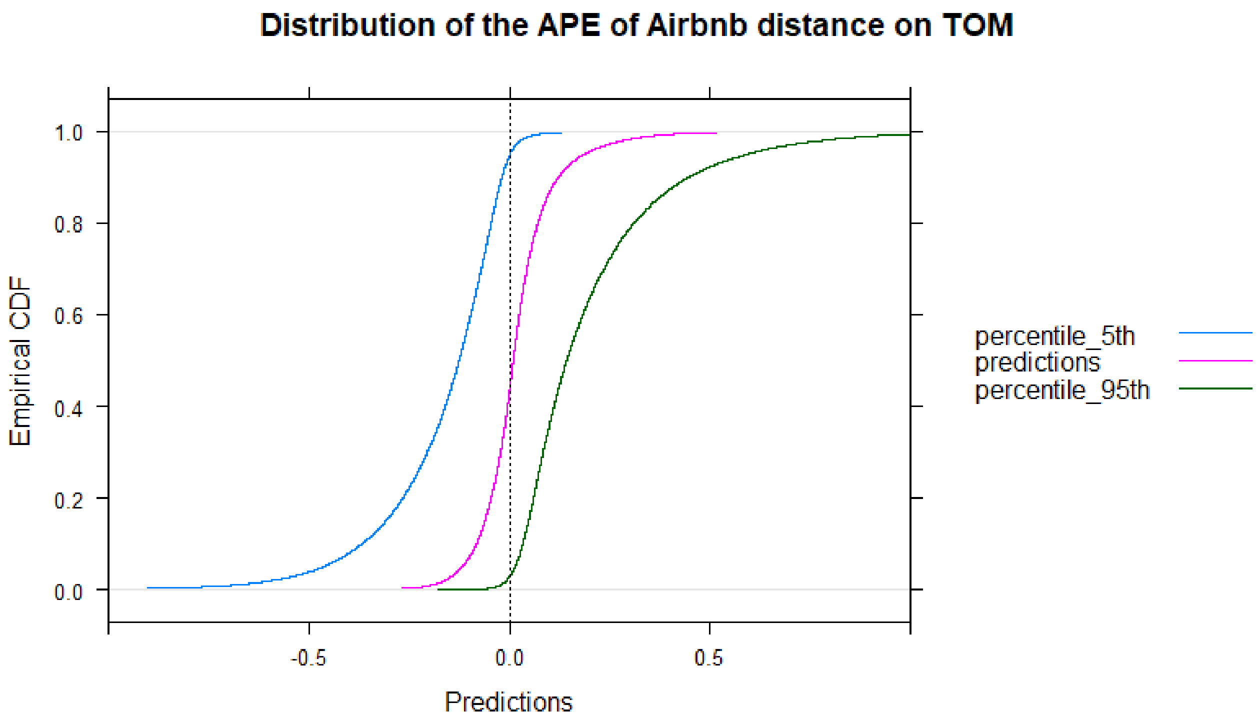


Figure 14: Random Forest treatment effect estimate for distance on time-on-market

The CDFs above show the point estimate predictions and prediction confidence intervals. The confidence intervals are very large and almost all of them contain the value 0.

6 Conclusion

While there is still lots of further research to do, the clear picture coming out of this analysis is that of heteregenous effects of Airbnb on house prices and to give a single estimate of the effect of Airbnb on house prices would be to misunderstand this completely. The other story coming out of this report is that by using the generalised random forest, which very flexibly accounts for a range of potential confounders, we find that the effect of Airbnb on house prices is quite small. In other words the Airbnb effect is not economically significant and does not warrant the attention and backlash Airbnb receives.

6.1 Policy Implications

Airbnb rental intuitively has positive and negative externalities, some of which are more concentrated such as noise and some which are more dispersed including spending in the local area. As there are positive externalities but the presence of Airbnb doesn't appear to uniformly lead to an increase in house prices this suggests that there are indeed negative externalities being imposed on the local area, reducing housing demand in certain areas.

When goods with negative externalities are supplied by individuals they tend to suffer from overconsumption. The negative externalities associated with Airbnb to the local neighbourhood may point toward the need for a local tax on such short-term accommodation. This local tax may be used to help facilitate more housing solutions for those people displaced by the rise in short-term accommodation in affected neighbourhoods. Given the potential positive externalities mentioned before, further reseach would need to be conducted before pursuing said local tax. This further research is outlined below.

As pointed out by Sheppard and Udell (2016), while Airbnb may have a positive impact on house prices, pursuing strategies that seek to lower house prices by reducing housing value are unlikely to be welfare-improving. An outright ban on home-sharing is, therefore, likely to be a bad policy.

6.2 Further Research

The variables capturing the effect of Airbnb are subject to measurement error which means the local partial effects estimated will be subject to attenuation bias and the true effects may be larger than those presented. It may be possible to identify an instrumental variable for the density of Airbnb within the local area. This would help address the issues associated with measurement error while also addressing the endogeneity challenges we have sought to address using Fixed and Time Effects.

The Analysis section began to explore the benefits of using the Generalised Random Forests approach. These results highlight the importance of having an estimation technique that is flexible enough to cater for these heterogenous local treatment effects which are typically assumed to be globally constant in more traditional panel data methodologies. The insight afforded by this flexible approach can better inform policy decisions rather than taking a blunt approach to policy making which can result in unintended consequences.

Further use could be made of the spatial data available. We investigated a preliminary high-level analysis using Moran' I: spatial correlation measure. The analysis, given the time constraints, was limited to the relationship between house price and the number of Airbnb listing. The more limited number of house price transactions in the period 2008-10 makes this form of simple analysis difficult and returned insignificant results over this period. However, we found statistically significant results for later years and identified that the house price correlation with first degree Queen neighbour is positive and that the strength of correlation increases over time. However, the second degree relationship is negative with increasing strength over time. An area for further research would be to include this form of neighbour relationships into the main model we have outlined above.

As discussed above, Airbnb may shift bargaining power from buyers to sellers and the resultant change in market microstructure. We conducted a preliminary investigation of how our two Airbnb variables affected time-on-market and didn't find a statistically significant relationship. However, further work would need to be conducted to fully resolve the issues with identification discussed. An area for future research will be to utilise an IV for Airbnb to identify an exogenous shift in the bargaining power. It may be possible to utilise the seasonal variation in the demand for short-time rental property. However, there is already seasonal variation in the housing market so a better understanding of the dynamics of the local housing market would be required to undertake this analysis.

References

[1] *Amsterdam to allow Airbnb rentals in city centre after court order*. Mar. 2021. URL: https://www.reuters.com/article/us-netherlands-airbnb-amsterdam-idUSKBN2B81NS.

[2] Susan Athey, Julie Tibshirani, and Stefan Wager. "Generalized random forests". eng. In: *The Annals of statistics* 47.2 (2019), pp. 1179–1203.

[3] Kyle Barron, Edward Kung, and Davide Proserpio. "The effect of home-sharing on house prices and rents: Evidence from Airbnb". In: *Marketing Science* 40.1 (2021), pp. 23–47.

[4] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[5] Michael W Browne. "Cross-validation methods". In: *Journal of mathematical psychology* 44.1 (2000), pp. 108–132.

[6] Victor Chernozhukov et al. *Double/debiased machine learning for treatment and structural parameters*. 2018.

[7] Thomas G Dietterich and Eun Bae Kong. *Machine learning bias, statistical bias, and statistical variance of decision tree algorithms*. Tech. rep. Citeseer, 1995.

[8] Jean Dubé and Diègo Legros. "A Spatiotemporal Solution for the Simultaneous Sale Price and Time-on-the-Market Problem". eng. In: *Real estate economics* 44.4 (2016), pp. 846–877. ISSN: 1080-8620.

[9] Liran Einav, Chiara Farronato, and Jonathan Levin. "Peer-to-Peer Markets". eng. In: *Annual review of economics* 8.1 (2016), pp. 615–635. ISSN: 1941-1383.

[10] Apostolos Filippas and John Horton. "The Tragedy of your Upstairs Neighbors: Is the Negative Externality of Airbnb Internalized?" In: June 2017, pp. 537–538. DOI: 10.1145/3033274.3085097.

[11] Miquel-Àngel Garcia-López et al. "Do short-term rental platforms affect housing markets? Evidence from Airbnb in Barcelona". In: *Journal of Urban Economics* 119 (2020), p. 103278.

[12] David Genesove and Christopher J. Mayer. "Equity and Time to Sale in the Real Estate Market". eng. In: *The American economic review* 87.3 (1997), pp. 255–269. ISSN: 0002-8282.

[13] Keren Horn and Mark Merante. "Is home sharing driving up rents? Evidence from Airbnb in Boston". In: *Journal of Housing Economics* 38 (2017), pp. 14–24. ISSN: 1051-1377. DOI: https://doi.org/10.1016/j.jhe.2017.08.002. URL: https://www.sciencedirect.com/science/article/pii/S1051137717300876.

[14] Claire Jones. *Brexit fuels Amsterdam property price boom*. July 2019. URL: https://www.ft.com/content/c060e162-98c1-11e9-8cfb-30c211dcd229.

[15] Kory Kroft and Devin Pope. "Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist". In: *Journal of Labor Economics* 32 (Apr. 2008). DOI: 10.1086/673374.

[16] Stephen Sheppard, Andrew Udell, et al. "Do Airbnb properties affect house prices". In: *Williams College Department of Economics Working Papers* 3.1 (2016), p. 43.

[17] Andrea Smith. *Amsterdam bans Airbnbs in certain areas of the city*. July 2020. URL: https://www.lonelyplanet.com/articles/amsterdam-bans-airbnb.