Transaction Price Effect of Number of Bidders in English Auctions in the Housing Market

Evidence from the Oslo apartment market

Dag Martin Sundelius

Master of Philosophy in Economics

UNIVERISTY OF OSLO

May 2017

Transaction Price Effect of Number of Bidders in English Auctions in the Housing Market

Dag Martin Sundelius

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

This thesis contributes to the literature of the micro-structure of housing transactions, by studying a specific type of housing transactions, the bidding war. It investigates the effect of number of bidders on the final transaction price, in English auctions for housing. From auction theory, the transaction price in an English auction equals the second highest valuation amongst the bidders. Therefore it should increase in the number of bidders if the valuation is heterogeneous amongst bidders. The thesis analyses the question at hand by utilizing a data-set detailing bidding rounds and unit attributes for apartments in Oslo. It employs the asking price, appraiser's valuation as well as a constructed hedonic pricing model including text-search variables, as baseline estimates for apartment value. It finds that increased number of bidders is associated with a higher final selling price, when the effect is measured as the spread from these baseline values. When the coefficients are standardized, the magnitudes are similar for the asking price and appraiser's valuation spread, but is substantially lower for the hedonic spread. Furthermore, by utilizing data from apartments sold more than once within the data-set, the thesis finds that number of bidders correlate over transactions for the same apartment. This indicates that the expected arrival rate of bidders is unit specific. Moreover employing apartment specific effects set-up, it finds that the estimated coefficient for number of bidders on the asking price spread, is positive and statistically significant when controlling omitted variables that are unit specific, further solidifying the findings.

Acknowledgements

I would like to thank my supervisor Erling Røed Larsen, for invaluable support and guidance. Furthermore I would like to thank Eiendomsverdi, in which I am part time employee, for suppling the data-set. Among my colleagues a special thanks is warranted to Andreas Jensen for discussion and proof-reading, as well as Enok Vatnar for help with the data-set.

The regressions and estimates in this thesis are performed in STATA, with additional figures created in Tableau. Do-files are available upon request. Any errors or inaccuracies, are solely my responsibility.

Contents

1 Introduction ... 1
2 Literature ... 4
3 Theoretical framework ... 7
4 Data and empirical methods ... 10
  4.1 Data ... 10
  4.2 Institutional description of the English auction in the Norwegian Housing market ... 11
  4.3 Summary statistics of bidding rounds ... 12
  4.4 Empirical methods ... 14
    4.4.1 Hedonic theory and method implications ... 15
    4.4.2 Hedonic model selection ... 17
  4.5 Data driven Challenges ... 20
5 Regression results ... 22
  5.1 Asking price spread ... 22
  5.2 Appraiser's valuation spread ... 23
  5.3 Hedonic predicted prices spread ... 24
  5.4 Comparison of coefficients ... 25
6 Discussion and extensions ... 27
  6.1 Alternative approaches ... 29
  6.2 Stochastic number of bidders ... 30
  6.3 Apartment fixed effects ... 32
7 Robustness check ... 34
8 Conclusion ... 35
References ... 36
Appendix ... 40
  A.1 Regression output from asking price and appraiser's valuation model ... 40
  A.2 Selection of functional form and regressors in the hedonic model ... 42
  A.3 Test for overfitting when using zip code as location dummy ... 48
  A.4 Description of variables ... 48

List of figures

1. Figure 1: Average monthly difference between final selling price and the asking price, as percentage of the asking price, for apartments in Oslo…………………………………………………………..2
2. Figure 2: Probability of achieving high price for different values of parameters N and Prob(B_g)…….9
3. Figure 3: Heat map where city districts with the highest average number of bidders have the darkest colors………………………………………………………………………………………………… 13

List of tables

1. Table 1: Distribution of transactions over years……………………………………………………...10
2. Table 2: Bidding rounds by city district……………..……………………………………………..12
3. Table 3: Bidding rounds by month…………………………………………………………………...12
4. Table 4: Regression asking price spread……………………………………………………………...22
5. Table 5: Regression appraisers' valuation spread……………………………………………….…23
6. Table 6: Performance of hedonic model……………………………………………………………...24
7. Table 7: Regression Hedonic spread…………………………………………………………………25
8. Table 8: Regression results: comparison of coefficients ……………………………………………….26
9. Table 9: Regression results stochastic number of bidders………………………………………………31
10. Table 10: Regression results, apartment fixed effects…………………………………………….…….33
11. Table 11: Regression output robustness check………………………………………………………….34
12. Table 12: Regression output for section 5.1……………………………………………………………40
13. Table 13: Regression output for section 5.2……………………………………………………….……41
14. Table 14: Functional form specification………………………………………………………………...42
15. Table 15: Selecting regressors from proxy group……………………………………………………….45
16. Table 16: Measures of fit for hedonic model with zip-code as location variable……………………….48
17. Table 17: Description of variables……………………………………………………………………….48

1 Introduction

Housing transactions have by economists traditionally been viewed as a matching and bargaining problem. In this framework asking prices induce search because they work as a ceiling for the final selling price, in that the seller must accept an offer from a potential buyer at that price. Furthermore, the final selling price is a result from bargaining between one potential buyer and the seller. (Han and Strange, 2013). However, some markets work in a different fashion, such as the current Norwegian housing market. Transactions in this market are often characterized by many bidders competing against each other in bidding wars and the final selling price being above the asking price. This begs the question exactly of how the interactions are between bidding wars and market outcomes. Or more precisely, does the number of bidders in a real estate auction increase the final selling price?

The effects of bidding wars could have several policy implications. Firstly, it should be of great interest to policy makers when regulating housing auctions, as information on the effect it has on final selling price could serve as foundation for policy decisions. Secondly, disentangling the final selling price on dwelling attributes and bidding round characteristics, should be of great significance for agents that are in the business of valuating real estate, such as banks setting loan-to-value ratios.

This thesis studies the bidding war phenomena, and examines apartment transactions in Oslo. More specifically it studies the effect of extra entrants into bidding rounds on the final selling price. The method utilized is to apply three different baselines for the value of an apartment: the asking price, the appraiser valuation and predicted prices from a hedonic model. Then estimating the effect of the number of bidders on spread over these baselines.

In Oslo, as of April 2017, prices for apartments are 58.9 percent higher than they were 5 years ago (Eiendom Norge, 2017). There has not been a month with negative average sell-ask spread, meaning the percentage difference between final selling price and the asking price, since March 2009.

Figure 1: Average monthly difference between final selling price and the asking price, as percentage of the asking price, for apartments in Oslo (Eiendomsverdi)

Although there is some previous work on bidding wars, empirical work on the effect of bidding wars on prices are rare, this might to some extent be due to historical rarity of the phenomenon, as well as differences in how housing transactions are conducted across countries. Another issue is that data on the bidding rounds itself are very rare. I have obtained a unique dataset detailing bidding rounds for apartments in Oslo, with both characteristics of the auction as well as attributes of the apartments themselves. This allows analysis that isolates the effect of the number of bidders.

In theory the number of bidders could affect the price in a number of ways. If we assume that buyers are totally homogenous in preferences and budget constraints, any number of bidders above 2 will have zero effect on the price. This is because in an English auction such as with Norwegian housing transactions, the final selling price equals the second highest bid plus the smallest possible value (McAfee and McMillan, 1987a). Genesove and Han (2011 unpublished) argue that if markets are very thick, and houses are homogenous, there would be very little difference in buyer's valuation of houses, given that people with different wealth and quality sensitivity search in different pricing segments. So within this framework the number of bidders should only affect the final price in thin markets. In thick markets, they argue that there is little to gain from further search from buyer and seller so the bargaining outcome with only one buyer, would be similar to the outcome where buyers compete.

These assumptions are strong. We can instead imagine that potential buyers only know their match utility after visiting the house, and this valuation in drawn from a random distribution. Then, as more and more buyers arrive at the house, the second highest valuation approaches the highest possible WTP(McAfee and McMillan, 1987a) and (Holt 1979). Thus the competition among bidders drives up the final selling price, this reasoning will be embellished upon in the theoretical frame work in section 3. Furthermore, there could be a behavioral effect where buyers in bidding war are caught in the heat of the auction. Either through reciprocity, or it could be that bidders update their valuation throughout the bidding war, when they see others bid.

The rest of the thesis is built up as follows:

Section 2, presents relevant literature of the micro structures of housing transactions, and previous empirical work conducted on bidding wars. Section 3 provides a theoretical framework within which to interpret the effect of number of bidders. Section 4 provides description and summary statistics of the data, as well as institutional description of Norwegian housing transactions, together with empirical methods description and description of challenges related to the data-set. Section 5 is the main empirical section, providing regression results for the effect of number of bidders on the asking price spread, appraiser's valuation spread and the hedonic spread. Section 6 provides discussion and extensions to the main empirical work. Section 7 provides robustness check for the main regressions, on identified observations. Section 8 gives concluding remarks to the thesis

2 Literature

The most simplistic models of the microstructure of housing transactions are one-sided search models, which study the problem of either seller or buyer, and take the other part as exogenous. The buyer’s problem is the existence of search cost and the fact that he cannot observe all attributes of house before he pays that cost e.g visits the house. It is first at that point he learns his match utility of the house. The seller’s problem, is setting the asking price. One example is that the arrival of potential buyers is random, and buyers must consider the asking price as take it or leave it. Here a lower asking price lowers the price, but it also decreases the expected time on market. Thus, the seller has a tradeoff between price and time on market. Another way to look at the seller’s problem, is that he too incurs search cost. In the way that he must pay search cost to attract potential buyers to his house, such as advertisement and staging the house. (Han and Strange, 2015, 820-824)

Less stylized models focus on the interconnection of buyers and seller, so called matching models. See for example Genesove and Han (2012). This is a so called random matching model, where the rate of contacts between buyers and sellers, is determined by the ratio of sellers and buyers, called market tightness. First after a contact is made, the buyers match utility is realized, drawn from a distribution. Whether a contact ends up being a match depends on the utility surplus generated, as a function of buyer and seller reservation values, and the match utility and thus the willingness-to-pay. After a match has been made, the buyer and the seller partake in a bargaining process. The process is described as a Nash bargaining problem over the distribution of the potential total surplus between the buyer and the seller. The outcome of this bargaining is a function of the total surplus and the relative bargaining power of the parties. One of the parts could choose to drop out of the bargaining process, however this depends on the cost of further search. (Han and Strange, 2015, 827)

Albrecht et. al (2015) constructs a directed search model with only limited commitment for the seller to the asking price. In the model, the seller is free to reject any offer below ask, but is committed to sell if one or more bids are received at or above ask. The model allows for competition among buyers above the listed asking price of a house, in the form of an auction. The paper reflects the scope of how transactions take place, with houses selling at, below or above ask. It also places emphasis on the role of the asking price, as a signaling tool of the seller, to attract buyers searching for houses. It also shed light on competition among sellers to attract buyers.

Han and Strange (2016) studies the role of the asking price in housing transactions. They analyze the part asking-price plays in attracting buyers to houses. With a lower asking price, a seller can promote more visits, thus increasing the expected number of buyers experiencing a high match utility with the house, thereby increasing the probability of a bidding war. However, there is some limit to this effect, as setting the asking price too low increases the likelihood of a potential buyer with a high valuation experiencing high competition in the bidding war. Hence a lower ask stops attracting more buyers at some point. They find support for their model in the empirics, indicating a lower asking price increases the number of bidders. They also find that the negative relationship between the asking price and the number of bidders is stronger in an atypical house.

Han and Strange (2014) studies on determinants of bidding wars in the real estate markets of Canada and the US. The authors do not have data on the number of bidders involved in a real estate transaction. Instead they define the occurrence of a bidding war as the transaction price of a house being higher than the asking price. They show that the occurrences of bidding wars are correlated with macroeconomic growth, and more specific housing booms, the latter both in prices and volume. They find that the frequency of bidding wars seem to be sticky, in the way that fall in housing prices do not cause the number of bidding wars in the market to fall down to its level before the boom. Furthermore the magnitude of the phenomena vary across cities. Less matching frictions also seem to impact the occurrence of bidding wars. As search cost for the buyer falls, there is a higher probability of having more potential buyers visiting the house and partaking in a bidding war. This is illustrated by bidding wars being correlated with internet advertisement of houses. As they point out it could explain some of the reason why the occurrence of bidding wars did not fall back together with housing prices, as increased internet use in the housing transaction process coincided with the drop in prices.

Genesove and Han (2011,unpublished) have access to a rare data set of the number of bidders. They have a survey from a large North American urban area, in which respondents report how many other bidders they competed against when they bought a house. Using maximum likelihood, they simultaneously estimate asking price, final price and the number of bidders. By doing so they try to avoid two issues. First, the inherent imprecise nature of hedonic regression models, caused by unobserved characteristics. Second, they argue that using the asking price introduces a bias, through the effect of the asking price on the number of bidders. They find that doubling the number of bidders increases the final on average, by 2.4 percent.

Sommervoll et. al (2015 unpublished) studies the effect of jump bids in housing bidding rounds. The use a dataset from a large Norwegian realtor, which contains information on the entire bidding round, including list of bids. Within this paper, they specifically control for the effect of number of bidders on the final price. They find a positive correlation between prices and the number of bidders.

Ashenfelter and Genesove (1992) finds that under an auction for condominium apartment units in New Jersey, apartments for which the sale fell through for some reason, sells for on average for 13% less than they were under the auction. The reselling was done through face to face bargaining.

This thesis adds to the literature of housing micro-structure by analyzing a specific real estate market type, the housing auction. Sommervoll et. al and Genesove and Han address the question of how number of bidders impact on the final selling price. However, an analysis with a specific hypothesis regarding number of bidders within the Norwegian framework should still be valuable, since there is likely to be institutional differences between the Norwegian and North American real estate markets.

3 Theoretical framework

This thesis is an empirical paper, however to understand the impact of the number of bidders on final selling price, this section develops a theoretical framework within which to interpret the effect.

Let the arrival of bidders to an housing auction be stochastic and Poisson(λ) distributed.

Then, let N be the number of bidders arriving at the house, the probability of n number of bidders arriving (1): $Prob(N = n) = \frac{\lambda^n e^{-\lambda}}{n!}$, in which

λ is the average arrival rate of bidders in a given time period. For simplicity, we assume that bidding round periods are equally long. From (1) we have:

(2): $E[N] = \lambda$

Furthermore, the bidder i has a willingness to pay for house h TP_{ih} , and it is described by the following function:

(3): $WTP_{ih}(M_{ih}, I_i) \quad \frac{\partial WTP_{ih}}{\partial M_{ih}} > 0 , \frac{\partial WTP_{ih}}{\partial I_i} > 0$,

in which M_{ih} is the match utility of bidder i for house h and I_i is the income of bidder i.

After arriving at the house the bidders observe their match utility M_{ih}, either M_g for a good match or M_b for a bad match, with $M_g > M_b$. Furthermore i may have either high or low income with I_H or I_L. We assume that both bidders with I_H or I_L search for all apartments, and has no preferences for apartments ex ante. The bidders can be categorized into two types $B = B_g \; and \; B = B_b$. So that N_j is the number of type $B_j, j = g, b$, $N_g + N_b = N$. Type B is characterized by the following equations

(4): $B = \begin{cases} B_g , M_{ih} = M_g \; and \; I_i = I_H \\ B_b, \; otherwise \end{cases}$

(5): $WTP_{B_g} > WTP_{B_b}$

Furthermore, after N bidders have arrived there is a probability that each bidder is of type $B_g or \; B_b$:

(6): $Prob(B_g) + Prob(B_b) = 1$

We assume that $N_g \sim Binomial\left(N, Prob(B_g)\right)$ with

(7): $Prob(N_g = n_g) = \binom{N}{n_g} Prob(B_g)^{n_g}(1 - Prob(B_g))^{N-n_g}$

(8): $E[N_g] = E[N]Prob(B_g) = \lambda Prob(B_g)$

Let the transaction price of a house determined by arrival and composition of the types of bidders. Competition among bidders will assure that $P = WTP_{N-1} + \epsilon$ when WTP is sorted in ascending order amongst the bidders, ϵ is the smallest possible increment with which to raise a bid.

As a simplification without loss of generality we assume that a house can either sell for a high price P_{high} or a low price P_{low}, or not sell at all, and that:

(9): $P = \begin{cases} P_{high}, N_g \geq 2 \\ P_{low}, N_g < 2 \end{cases}$

If N=1 there is a bargaining game between seller and buyer, we assume P_{low} as a final outcome independently of the type of bidder. If N=0 the house is taken off the market.

(10): $Prob(P_{high}) = 0 \; when \; N = \{0,1\}$

Combining (1) and (10) gives (11): $Prob(P_{high}) = 0: \lambda e^{-\lambda} + e^{-\lambda}$

In general:

(12): $Prob(P_{high}) = 1 - Prob(N_g = 1 \cup N_g = 0)$

thus,

(13): $Prob(P_{high}) = 1 - \binom{N}{1}Prob(B_g)^1(1 - Prob(B_g)^{N-1}) - \binom{N}{0}Prob(B_g)^0(1 - Prob(B_g))^{N-0}$

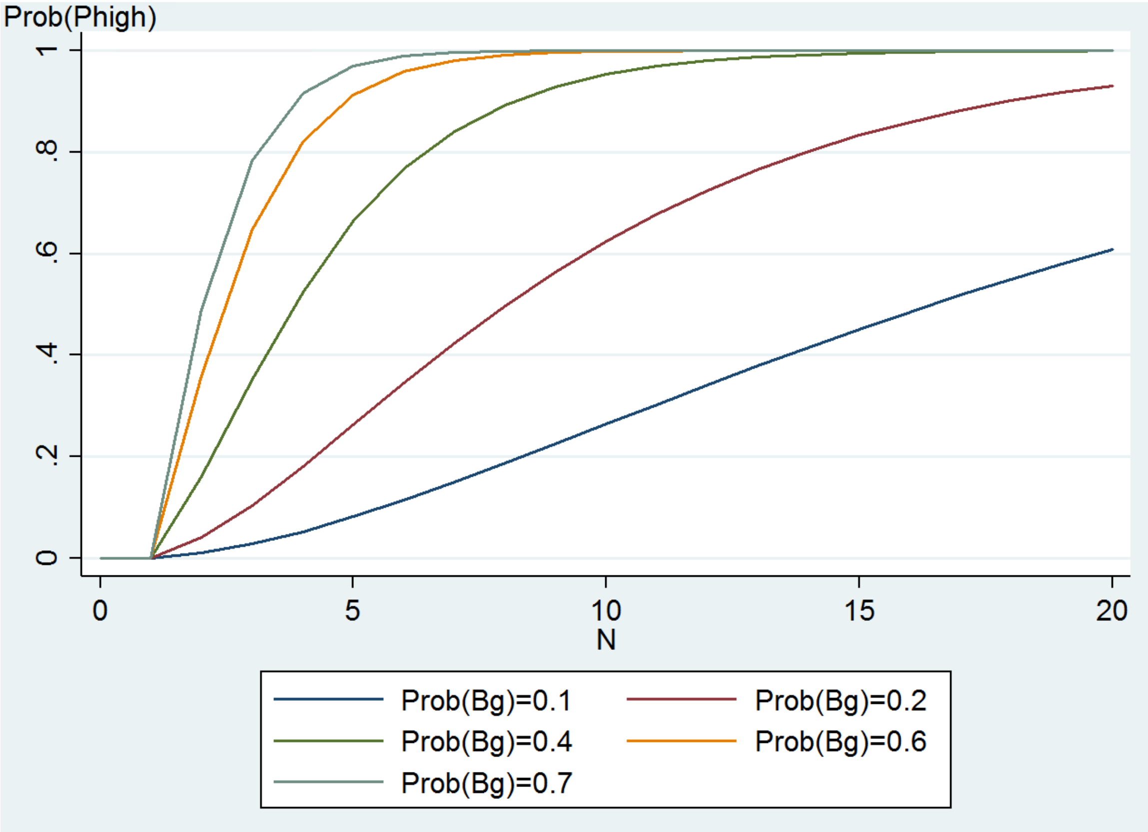
(14): $Prob(P_{high}) = 1 - NProb(B_g)(1 - Prob(B_g))^{N-1} - (1 - Prob(B_g))^N$

Thus, we obtain (15) for the properties of $Prob(P_{high})$ as a function of N

(15): $Prob(P_{high})_{(N)} - Prob(P_{high})_{(N-1)} > 0, \; N \geq 2$

I have plotted (14) for different values of the parameter *Prob(B_g)* in figure 2 from this it is clear that (15) holds. And illustrates that within this framework the probability of achieving a high final selling price is indeed increasing in the number of bidders.

Fig 2: Probability of achieving high price for different values of parameters N and Prob(B_g)



The presented framework above, assumes that the number of bidders are drawn from a totally stochastic distribution, this assumption is maintained throughout the empirical sections. The extent to which the assumption holds will be discussed further in section 6.

4 Data and empirical methods

4.1 Data

The data set used in this thesis contains data on the bidding round for transactions of apartments in Oslo, it covers 54 139 bids from 5 666 sales. The data is provided by Eiendomsverdi AS, a commercial company gathering and developing Norwegian real estate statistics. There are data from 5 realtors. The data set spans from January 2013 to January 2017, the distribution of transactions over years is shown in Table 1.

Table 1: Distribution of transactions over years

Year	2013	2014	2015	2016	2017
Share of transactions	3 %	20 %	43 %	34 %	<1%

In the same period a total of 67 174 apartments where sold in Oslo (Eiendomsverdi AS). The data includes information on the bidding rounds themselves such as the number of bidders and listed interests. Furthermore the data set contains the asking price of the listing, along with attributes of the apartment, such as size (livingarea Norwegian “P-rom”), location and number of floors etc. The data include the header text of the advert associated with the transaction. This is used to create dummy variables for whether the header mentions certain words, for example: Is the Norwegian word for a view mentioned?

Eiendomsverdi AS has also provided supplementary data. These data include appraiser’s valuation, which zip code belongs to which city district within Oslo, and how many dwellings of a zip code lies within a certain city district. Any direct info on city district domain is not included in the dataset from the realtors, thus to construct location variables, zip codes have to be translated to city districts. Furthermore Eiendomsverdi AS has supplied data for graphs and facts for the Oslo apartment market in this thesis, these have been cited as “(Eiendomsverdi AS”)

See the appendix for a list and description of all variables.

4.2 Institutional description of the English auction in the Norwegian Housing market

Bidding rounds in the Norwegian housing market are a special form of an English auction. A traditional English auction is according to (McAfee and McMillian, 1987a) characterized by the following:

- The auction starts with the announcement of a suggested opening bid
- The standing price gets raised successively by entered bids
- At any point each bidder knows the current highest bid
- The auction closes when the highest bid has been left unchallenged for a given amount time
- At this point the item is sold to the bidder with the highest bid for a price equaling the value of that bid

The Norwegian housing market auctions differ slightly from the traditional English auction, some of the traits of the bidding round as described by Norwegian Association of Real Estate Agents, NEF (2016) are the following:

The auction is not held at an auction house, rather bids are sent in a written form to the realtor. This is typically executed either in person at the viewing, through sms or online. To be able to bid, a person needs to provide personal identification. Furthermore, a bid is legally binding from the moment it is communicated by the realtor to the seller, and cannot be withdrawn. However, as opposed to a typical English auction, a bid has an expiration deadline after which the bid is no longer binding. Within that period, the bid is binding unless the house is sold to another bidder or the bid has been declined by the seller. The length of this deadline varies, but it cannot be set earlier than 12:00 the first business day after the last advertised viewing. After this point, the period between the moment the bid is transmitted and the deadline, has to be sufficiently long, so that the realtor has time to inform the other parties. The bid remains binding even if it is surpassed by a higher bid. This is because the seller is not required to sell to the highest bidder. A sale to a lower bid can occur because of demands included in the bid, such as inclusion of appliances or a different takeover date. Furthermore, for dwellings in housing cooperatives, members sometimes have right of first refusal. This means that member have the right to purchase the unit at the price of the winning bid.

4.3 Summary statistics of bidding rounds

On average an apartment in Oslo has 15.24 people noted as interested in connection with a viewing. There are on average 3.10 bidders in a bidding round, bidding a total of 9.47 times.

Table 2: Bidding rounds by city district

City District	*Avg. Number of bidder*	*Avg. Number of interested*	*Avg. Number of bids*
Alna	3.27	12.36	10.39
Bjerke	3.07	13.33	10.16
Frogner	2.87	16.25	9.00
Gamle Oslo	3.59	18.04	10.60
Grorud	3.52	11.62	12.14
Grünerløkka	3.33	19.60	9.86
Nordre Aker	2.95	14.87	10.15
Nordstrand	2.42	10.47	7.35
Sagene	3.47	21.19	10.06
Sentrum	3.22	13.41	9.26
St. Hanshaugen	3.14	17.69	9.44
Stovner	3.33	11.86	12.49
Søndre Nordstrand	2.38	8.05	7.59
Ullern	2.42	13.19	7.92
Vestre Aker	2.22	10.34	7.31
Østensjø	3.39	12.21	10.11

Table 3: Bidding rounds by month

Month	*Avg. Number of bidder*	*Avg. Number of interested*	*Avg. Number of bids*
January	3.21	14.61	9.81
February	2.94	13.52	9.16
March	2.85	13.46	8.85
April	3.09	15.71	9.97
May	3.03	14.35	9.43
June	2.92	13.95	9.13
July	3.78	23.44	9.82
August	3.17	17.31	10.08
September	2.96	15.51	9.23
October	3.01	15.12	9.81
November	3.09	15.44	9.61

Figure 3: Heat map where city districts with the highest average number of bidders have the darkest colors

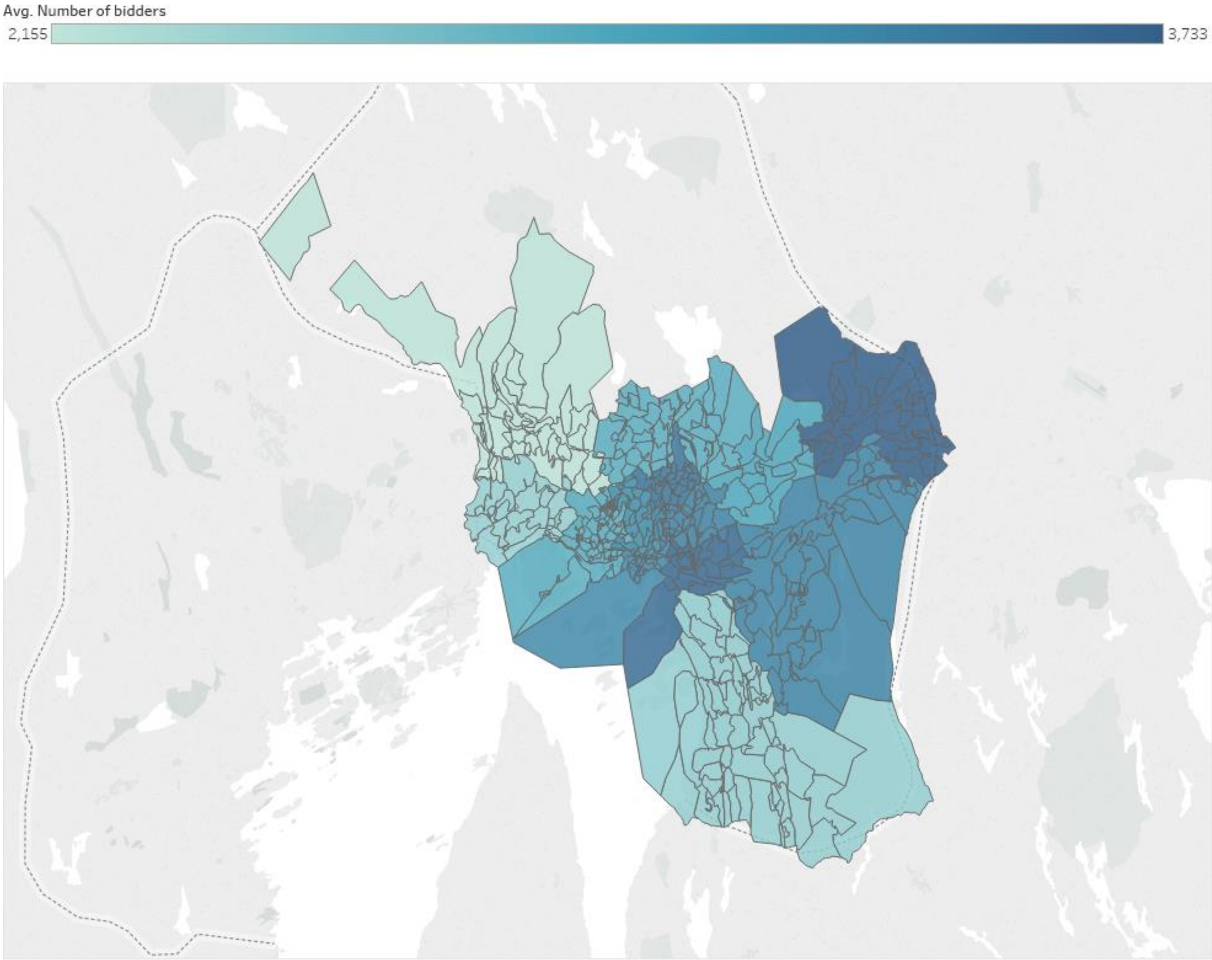


Table 2 shows the seasonality of bidding round intesity. July is the month where bidding rounds are the most intense symbolized by the high number of bidders and number of interested, with the other months having somewhat evenly distributed stats.

Table 3 shows the geographic distribution of bidding round characteristics, broken down on city districts, with the district of Marka not included because of no observations of apartment trasanctions. With Gamle Oslo, Grorud and Sagene being the city districts with the most intense bidding rounds, while Vestre Aker, Søndre Nordstrand and Ullern being on the opposite end of the scale.

Figure 3 shows the mean number of bidders as shown in table 2, in the form of a heat map. It is possible to detect a center periphery pattern, where the city center having the most intense bidding rounds. However, Grorud in the top right corner, contradicts this pattern.

4.4 Empirical methods

The research question of this thesis is the following: Does the number of bidders affect the final selling price? In order to answer this we turn to a regression analysis of the Oslo apartment market. To measure the impact of the number of bidder, we need a baseline with which to gauge "high" or "low" transaction prices. As with many other empirical issues, with real estate transaction one does not observe the counterfactual. We know what the apartment was sold for, but the price for the same apartment at the same time but with different bidding round characteristics is unknown. This thesis utilize three such baseline values, to estimate the effect of the number of bidders on the final selling price:

- The asking price
- The appraisers' valuation
- Hedonic predicted price

My empirical technique is to employ regression analysis, with following form:

(16): $S_i = \beta_0 + \beta_1 N_i + \beta_2 \delta_{1_i} \dots \dots + \beta_n \delta_{j_i} + u_i$

(17): $S_i = \frac{Price_i - V_i}{V_i}$

V_i is the asking price, appraiser's valuation or the hedonic predicted price. V_i also includes the common debt, as it constitutes part of the financial transaction. N is the number of bidders, δ is a vector of control variables and u_i is the residuals of the model, assumed $u \sim N(0, \sigma^2)$. The standard errors reported in this thesis unless stated otherwise, are the White heteroskedastic robust standard errors. An important note is that when the dependent variable is a spread, it could vary both in V and the Price, having a large asking price spread for example, could be both because a low asking price or a high final selling price.

There are several known issues with the asking price, which are identified in the literature section, section 6 discusses these further. Therefor the appraiser's valuation is employed to have another estimate that does not suffer from the issues of the asking price. It is set by an authorized appraiser who inspects the unit, and the valuation reflects the condition, features of the apartment and general market characteristics. It is constituted to be the appraiser's best estimate for a market price. Appraisers should not have any direct incentives connected to the final selling price. As they do not earn commission based on the price, nor do they benefit in a low time on market. The appraiser's valuation should at face value be independent of the number of bidders.

4.4.1 Hedonic theory and method implications

The third method for implementing a baseline value for an apartment, is a hedonic pricing model. The advantage of this approach is that it avoids human biases in predicting the price. However, as is well known the hedonic model introduces challenges of omitted variables and functional form. I will explain how I deal with these challenges below.

Hedonic pricing theory attempts to infer the unobserved and implicit prices of attributes from observing the total price of a good and the attributes which makes up this good. Thus, it is a revealed preferences method. More formally, to use the notation and argumentation of Rosen(1971): A good is represented as a vector of coordinates (18): $z=(z_1,z_2,\ldots\ldots,z_n)$ in which z_i is a measure of ith attribute of the good. Different consumers may consume goods with different z. Thus a sale of the good is the sale of a bundle of attributes. The price of these sales obtained in equilibrium is then (19): $p(z)=p(z_1,z_2,\ldots\ldots,z_n)$, and the implicit price of an attribute is then (20): $\frac{\partial p}{\partial z_i} = p_i$ where $i = 1, \ldots, n$.

The hedonic approach tackles the problem of heterogeneous goods, by decomposition. Therefore, even though the goods do not have a common price, each attribute is assumed to have one. Analysis of demand structures is possible through using these prices. (Sheppard 1999).

Hedonic theory is commonly applied to the analysis of housing markets. This is because real estate is in large part extremely heterogeneous. It is heterogeneous both vertically, meaning difference in easily observable characteristics among units such as size etc, and horizontally, meaning differences in harder to observe attributes pre viewing, such as the view and noise etc. (Nenov et. al 2016).

The framework is used either to estimate prices for individual attributes, or to predict housing prices. This is typically done through multivariate regression where the final selling price is assumed to be a function of attributes $f(\mathbf{X}, \beta)$ where x is a matrix of individual attributes and β is a vector of estimated coefficients.

Hedonic regression analysis of housing markets, as mentioned involves challenges. Most of them it shares with general multivariate regression, and some from the fact that we are estimating market outcomes, which are functions of supply and demand.

Firstly, functional form is a typical issue because hedonic theory places few restrictions on it. If prediction is the goal, a criterion that measures which form best fits the data, might be optimal. However, doing so might come at the cost of causal interpretation of the coefficients in the hedonic price function (Cropper et.al 1988)

Secondly, a more general problem of model specification is the selection of regressors and how they are included the final model. The marginal value of an additional square meter for example, will differ by location, so including interaction term between size and location is essential. Also as a consequence of the spatial structure of housing, there might be spatial autocorrelation, this makes observations and thus errors not independent within submarkets (Can,1992). Moreover issues of collinearity between variables seems intrinsic in the housing market. As pointed out in Sheppard(1999), even though housing is heterogeneous, because of similarity in preferences and constraints in the form of technology and regulations, the variance of bundles of characteristics we observe is limited. This increases the probability of variables tending to move together.

Thirdly as prices are outcomes of demand and supply, estimating the hedonic demand equation is subject to simultaneous equation bias. The observed prices and quantities are intersections of demand and supply and therefore fitting an OLS line through these points estimate neither a demand nor a supply curve, because they are results of changes in both (Stock and Watson, 2015, 475)

Lastly as there almost infinite amount of attributes that determines the price of a house, many of which are unobservable to those who employ hedonic regression analysis. The model therefor intrinsically suffer from omitted variable bias.

4.4.2 Hedonic model selection

The goal of the model selection is, to the best of my ability, address the issues listed in the previous section, while building a model that estimates the effect of number of bidders on the final selling price.

The model design creates a baseline estimate for the value of an apartment, to estimate the effect of number of bidders on final selling price. It will include control variables that might correlate with both number of bidders and the final selling price, to minimize omitted variable bias. However including irrelevant variables is likely to increase the variance of the estimated coefficient of interest (Verbeek,2012,p 63). Therefor there is a tradeoff that has to be kept in mind.

The data set is quite rich in terms of available variables for the hedonic model. I will start off with a fairly general model which is split into two parts, one with variables that stem from theory, and another part with variables that are proxies for attributes of the apartment, that typically is unobserved in datasets used for hedonic regressions. The inclusion of the latter category is an attempt to catch some of the mentioned horizontal heterogeneity. I will then stepwise drop variables that have a t- values < |1,96| from the proxy category, while keeping variables in the former even if they have low t-values. More precisely, the variable with the lowest t-value while it has t-value <|1,96| from the proxy category will be dropped. Then the regression will be run again dropping the least significant variable. This is continued until there are no more “proxy variables” that has t-value<|1,96|. There are many critics of such stepwise regression, see for example Thompson (1995). The method will by design lead to variables that has t-values ≥|1,96|. So their true p-values differs from the reported once, as the p-values do not reflect the variable selection process (Verbeek,2012,p 76). Furthermore the final model will tend to fit better in sample than out of sample. It is however difficult to determine a priori whether or not variables in the proxy group determines the final selling price. Therefor despite these issues the method will be employed in the model selection. However the stepwise regression will only be used on the proxy group. Moreover, the coefficients of the hedonic model are not meant to be interpreted causally.

Regarding the issue of simultaneity, the hedonic model regression with housing attributes the coefficients is not meant to be causal. They are included to predict the final selling prices, to provide a baseline value for apartments. Furthermore the number of bidders are obviously not an attribute, that are demanded and supplied, but a feature of the bidding round itself.

The unrestricted hedonic model before functional form specification is of the form:

$$(21)\colon P_i = \alpha_0 + \alpha_1 NB_i + \alpha_2 F_I + \alpha_3 SI_i + \alpha_4 BY_i + \alpha_5 G_i + \alpha_6 CO_i + \theta D_i + \gamma DI_i + \delta SI_i XDI_i + \beta_1 FI_i + \beta_2 V_i + \beta_3 NR_i + \beta_4 R_I + \beta_5 SU_i + \beta_6 B_i + \beta_7 HR_I + \beta_8 ER_i + \tau CD_i + \epsilon_i$$

NB is the number of bedrooms of apartment i. F is the floor of apartment i. SI the size of apartment i. BY the build year of apartment. G is a dummy that indicates whether an apartment is sold with a parking spot in a garage. CO is a dummy for housing cooperative. D is a vector of dummies for time fixed effects for transaction date with month dummies and year dummies. DI is a vector of dummies location for fixed effects with city district dummies. SIXDI is the interaction term between location and size, allowing the effect of size to vary between city districts. FI,V, NR, R, B and SU are dummies for whether the header mentions the Norwegian word for fireplace, view, needs renovation, renovated, balcony and sunny. HR is the heating rating. ER is the energy rating. CD is the commondebt of apartment i[1]. ϵ_i is the residual of the regression with assumed $\epsilon_i \sim N(0, \sigma^2)$. CD, NB, F, SI, BY, G, CO, D, DI, SIXDI, are in the group which will be kept regardless of their statistical significance, however

[1] Although common debt is a part of the financial transaction, it is not given that it is perfectly reflected in in the final selling price. See for example Theisen(2016) for a discussion regarding the reflection of common debt in the final selling price.

they might be subject to functional form respecification. FI, V, BR, R, SU, B, HR, ER makes up the “proxy”group that might be dropped depending on their significance.

An issue with the location variables is which zones to employ for geographical positioning. The original dataset included, as earlier mentioned, zip codes not city districts. The reason why city district has been chosen over zip codes is that fitting the model with such fine grid of location dummies (there are 450 zip codes in Oslo) risks overfitting issues. The model would try to explain much of what is essentially just noise, and there would be large goodness of fit differences between out of and in sample prediction. I have tested whether this is an issue in the appendix.

Given the variables included in the unrestricted model, the next issue is the functional form. There is several ways this could be done. One way of addressing this is using the Box-Cox transformation first developed by Box and Cox(1964), in which the practioner transforms the dependent variable here P as (22): $P^{(\lambda)} = \frac{p^{\lambda}-1}{\lambda}$ then determining λ, and there by the functional form of P, through maximum likelihood estimation. As a consequence of the functional form of the dependent variable in the main regression, the final estimate will already be in the form of a percentage point effect. I have therefore kept the left hand side of the hedonic model linear[2]. As the unrestricted model is quite large, there are many alternative right hand side functional forms. To test all of them would use up a substantial amount of the degrees of freedom. An alternative is the Ramsay RESET test, where the analyst test whether powers of the fitted values help explain y, in an auxiliary regression (Verbeek,2012,p 71). However a large issue here is that the test also catches omitted variable bias, as a hedonic regression model typically suffers from this. Therefore the choice is to identify candidate variables before the running the model and include powers of them, and check if their powers are significant, then stepwise dropping the highest power of the variables if they have t-value <|1,96|. These variables are: NB, S, BY and F. CD is excluded from this group even though it is not a categorical or dummy variable, as the final selling price variable includes the common debt of the apartment, and therefore the marginal effect should be linear. The regression outputs leading up to the final model are in the appendix.

[2] Another argument for keeping the left hand side linear, is Jensen’s inequality. As we are predicting linear prices, then as $E[y_i|x_i] \geq \exp\{E[\log y_i|x_i]\}$, taking the exponential of the predicted logarithmic prices, is not a good predictor for the prices (Verbeek,2012,p 61)

4.5 Data driven challenges

In its origin, the data is generated from manual input by the realtors during and after the bidding round. This might potentially lead to errors in input. I do not have confirmation that the data-set is quality-controlled in its origin. I have therefore manually inspected extreme values of the variables. There are a couple of obvious errors in inputs. I have dropped observations where there are without a shadow of a doubt errors in input. There are 2 observations that are dropped because of this reason. One case where the appraiser's valuation is 1 950 NOK, with salesprice equaling 2 250 000 NOK and asking price 1 950 000 NOK. The other case is an observation with buildyear equaling 19 652. It is important to underline that these two observations are not dropped only because they are extreme, but because they are nonsensical and impossible.

Through inspection of the bidding rounds, I have identified 373 transaction where a bidder has raised his bid without his standing bid being surpassed by another bidder, while the standing was above the asking price. If the asking price is the true reservation price of the seller, this should not happen. However, there could be several reasons why the seller refuses a bid above asking price, see section 4.2. I have kept these observations in the data, however I cannot reject that the number of bidder variable is subject to errors in input for these observations. The number of bidder variable, is constructed as the count of a unique bidder specific identification number. If for these observations the identifiers are incorrectly applied, the number of bidders variable is smaller than its true value. Therefore, I have conducted a robustness check for these observations in section 7.

I have also dropped observations which from the header text it is clear that are not apartments. These are 36 parking places, 4 apartment building sold as a whole and 93 transaction of realtor contracts.

Furthermore, there are several observations with missing variables, observations where variables in the unrestricted hedonic model are missing, these observations have been dropped. The appraiser's valuation is also missing from a part of the data-set this leads to the final number of observations: 4 509 for the hedonic and asking price method, and 3 293 for the appraiser's valuation method.

As a note on the placement of apartment into city districts, zip codes are placed in to the city district which contains most of its dwellings. This is approximate only as there are occasions where the number of dwellings are split quite evenly among city districts.

A note on the creation of dummy variables from the header text: The dummies are only proxies of the attributes, because they do not necessarily correlate perfectly with the actual attribute. There might be a fireplace located in the apartment even though it is not mentioned in the header. On the other side, there might be some “sugar-coating” in the header text, a view might constitute many things, however this would likely surface as the variables being statistically insignificant the regressions at hand.

5 Regression results

5.1 Asking price spread

The asking price baseline regression uses the model: (23): $S_{a,i} = \beta_0 + \beta_1 N_i + \beta_2 \delta_{1_i} \ldots\ldots + \beta_n \delta_{j_i} + u_i$

Here (24): $S_{a,i} = \frac{Price_I - V_i}{V_i}$, $V_i = askingprice_i$

u_i is the residuals of the model, assumed $u \sim N(0, \sigma^2)$.

The control variables δ are size, city district, and dummies for transaction date, these are included to correct for the fact that high S a due to for example relative low asking price and high number of bidders might be features of certain sub-markets, without necessarily involving a causal relation between the two.

Table 4: Regression asking price spread

	(1) S_a	(2) S_a	(3) S_a	(4) S_a
Number of bidders	0.0246***	0.0238***	0.0233***	0.0221***
	(0.0005)	(0.0005)	(0.0005)	(0.0004)
Controlled for Size	No	Yes	Yes	Yes
City district FE	No	No	Yes	Yes
Month and Year FE	No	No	No	Yes
Constant	-0.0005	0.0367***	0.0101	-0.0157
	(0.0016)	(0.0035)	(0.0139)	(0.0147)
Observations	4509	4509	4509	4509
Adjusted R^2	0.47	0.48	0.50	0.55

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

See appendix for full regression output

Specification 4 in table 4 controls for size together with time and location specific effects. The results indicate that the coefficient estimate for the number of bidders is robust against possible confounders. One extra bidder is associated with 2.21 percentage points higher asking price spread holding size, location and time constant, which is significant even at the 1 percentage level. As a reference point the mean sell-ask spread is 7.8 %. Quite interestingly, the Adj. R^2 of 0.47 in the specification 1, indicates a large part of the sample variance of the asking price spread is explained by the number of bidders.

5.2 Appraiser's valuation spread

The appraiser's valuation baseline regression echoes the previous section with the model

(25) $S_{v,i} = \beta_0 + \beta_1 N_i + \beta_2 \delta_{1_i} \ldots\ldots + \beta_n \delta_{j_i} + u_i$

Where (26): $S_{v,i} = \frac{Price_I - V_i}{V_i}$, $V_i = appraisersvaluation_i$

u_i is the residuals of the model, assumed $u \sim N(0, \sigma^2)$.

The control variables δ are size, city district, and dummies for transaction date

Table 5: Regression appraiser's valuation spread

	(1) S_v	(2) S_v	(3) S_v	(4) S_v
Number of bidders	0.0235***	0.0228***	0.0222***	0.0212***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Controlling for Size	No	Yes	Yes	Yes
City district FE	No	No	Yes	Yes
Month and Year FE	No	No	No	Yes
Constant	-0.0122***	0.0224***	-0.0083	-0.0490*
	(0.0020)	(0.0045)	(0.0203)	(0.0212)
Observations	3293	3293	3293	3293
Adjusted R^2	0.38	0.40	0.41	0.47

Robust standard errors in parentheses

See appendix for full regression output

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Regression results indicate as with the asking-price spread, that the estimate of the coefficient of number of bidders is robust to potential confounders. Controlling for size and time and location, increasing the number of bidders in a bidding round with 1 is associated with the finale selling price being an extra 2.1 percentage points above the appraiser's valuation, this is significant at the 1 percentage level. As a reference, the mean difference between price and appraiser's valuation in the data is 5.9 percentage. We note that the Adj. R^2 of 0.38 in the uncontrolled model is similar to the same metric for asking price spread, and quite large.

We note that the coefficient for N in the fully controlled specification is close to the 0.0221 found for the sell-ask gap in the previous segment. The denominators are of course different, but asking price and valuation tend be quite close (sample corr=0.95).

5.3 Hedonic spread

The hedonic predicted prices baseline regression uses the model: (27) $S_{p,i} = \beta_0 + \beta_1 N_i + \beta_2 \delta_{1_i} \ldots\ldots + \beta_n \delta_{j_i} + u_i$

Where (28): $S_{p,i} = \frac{Price_I - V_i}{V_i}$, $V_i = predictedprice_i$

u_i is the residuals of the model, assumed $u \sim N(0, \sigma^2)$. The δ control variables are not included in this method, since they are already included in the predicted prices.

The hedonic model used to predict the price is the following after functional form and regressors selection have been executed (regression output for model specification, see appendix):

$$(29): V_i = \alpha_0 + \alpha_1 NB_i + \alpha_2 NB_i^2 + \alpha_3 F_I + \alpha_4 SI_i + \alpha_5 SI_i^2 + \alpha_6 BY_i + \alpha_7 BY_i^2 + \alpha_7 BY_i^3 + \alpha_6 G_i + \alpha_7 CO_i + \theta D_i + \gamma DI_i + \delta SI_i XDI_i + \beta_1 FI_i + \beta_2 V_i + \beta_3 NR_i + \beta_4 R_i + \tau CD_i$$

This the model 5 in table 14 in the appendix. To test how well the model performs I split the sample in two, where 70% of the sample is used to estimate the coefficients and then test how well it performs in explaining the final price in the remaining 30% out of sample. The sample split is done by using a random number generator.

Table 6: Performance of hedonic model on observed prices

	R^2	Hit-rate*
In sample	0.85	89.1%
Out of sample	0.79	88.9%

*hit-rate(share of sales prices within 20% of predicted prices)

As we can see there is very little difference in performance within and out of sample, any symptoms of overfitting cannot be detected from this table. Note that the sampling is conducted once and there be might some variance in the sample estimates, due to the composition of the groups.

We now look at how the number of bidders effects the gap between the predicted prices from the hedonic model and the final selling price. Note that this estimation uses the whole sample, both to predict the prices and the effect of the number of bidders.

Table 7: Regression Hedonic spread

	S_p
Number of bidders	0.0067***
	(0.0008)
Constant	-0.0175***
	(0.0033)
Observations	4509
Adjusted R^2	0.01

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As we can see, one extra bidder is associated with a 0.7 percentage point higher final selling price when normalized to the predicted price from the hedonic regression model. This is significant even at 1 percentage level. Note that the adjusted R^2 is the smaller than in the previous methods.

5.4 Comparison of coefficients

The coefficients of the number of bidders are not directly comparable, since the unit of S differ between methods, $S_a = \frac{Price-askingprice}{askingprice}$, $S_v = \frac{Price-appraisersvaluation}{appraisersvaluation}$,

$S_p = \frac{Price-predictedprices}{predictedprices}$ To compare the estimates, one needs to standardize the coefficients. This is done in the following way, (30): $S_{stand.} = \frac{S-\mu_S}{\sigma_S}$

Where μ_S is sample averge of S and σ_S is standard deviation of S. We rerun the regressions for the three methods, with the uncontrolled regression specifications.

Table 8: Regression results: comparison of coefficients

	(1) Stand. S_a	(2) Stand. S_v	(3) Stand. S_p
Number of bidders	0.2956^{***}	0.2900^{***}	0.0512^{***}
	(0.0056)	(0.0072)	(0.0060)
Constant	-0.9407^{***}	-0.8870^{***}	-0.1629^{***}
	(0.0191)	(0.0246)	(0.0254)
Observations	4509	3293	4509
Adjusted R^2	0.47	0.38	0.01

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The coefficients now have the unit of standard deviations. For example, 1 additional bidder is associated with an increase 0.2956 standard deviation units, when using the asking price method. The regression outputs indicates that the appraiser's valuation and asking price methods gives very similar results, however the predicted prices method gives a substantially lower coefficients for the number of bidders. This perhaps indicates some endogeneity between the asking price, appraiser's valuation and the number of bidders. Note that the appraiser's valuation sample differs from the other methods.

6 Discussion and extensions

The regression results in section 5 indicate that increasing the number of bidders is associated with an increase in the final selling price, when measured as the deviation from the asking price, appraiser's valuation and the hedonic predicted price. If this indicate a causal relationship between the number of bidder and the final selling price full stop, there are several implications. A change in regulation regarding housing auctions, should be done while keeping in mind the price implications. For example, allowing shorter acceptance deadlines in the bidding rounds, may exclude bidders from entering the auction, and could potentially lower the final selling price. Furthermore, banks assessing risk in their mortgage-portfolio through LTV-ratios, should have great interest gathering data on housing auctions, to increase their precision in estimating the value of a house.

However, interpreting the findings further into a causal relationship between the number of bidders and final selling price full stop, is non trivial.

Firstly, the asking price is not an unbiased estimate of the final selling price of an apartment, the average asking price spread in the sample is 7.8 %. This is not a problem in our analysis if it is independent of the number of bidders. If the only channel number of bidders affects the asking price spread is the final selling price, these results would constitute strong evidence to support the hypothesis of bidding war effects. However, as pointed out in the literature section, asking price itself can have an effect on the number of bidders attracted to an apartment. If sellers and realtors think that setting a low asking price increases the number of bidders and thus the price, they have incentives to set a low asking price. It is important to keep in mind that setting an artificially low asking price is not necessarily done to mislead buyers. If prices in a market typically goes above asking price, buyer's expectation would also probably reflect this. This leads to a Nash-equilibrium where a seller who deviates, and sets their asking price equal to their reservation price, will attract fewer buyers. It might lead to sticky asking prices, both upwards and downwards. Sell-ask spread will remain high after booms and low and busts (Knight et. al, 1994). On the other side setting a low asking price could influence the valuation of the buyers downwards, if they use the asking price as a reference point. This is called the anchoring effect by behavioral economists, see for example

Northcraft and Neale(1987) for the anchoring effect of the listed price in real estate markets. In conclusion, the estimate of the impact of the number of bidders on the asking price spread, is both the effect an extra bidder has on the final selling price, but also the effect the asking price has on the number of bidders. The coefficient for number of bidders would then be biased in the asking price method if we interpret it as the effect on the final selling price alone.

Secondly, the number of bidders coefficient found from the appraiser’s valuation method is similar to the asking price method. This might indicate two things. Firstly, it could be that there are a very little endogenous effect from the asking price on the number of bidders, which would mean that the results are close to a causal estimate of the effect of the number of bidders on the final selling price. On the other hand, it could be that appraiser’s valuation too is endogenous, which would leave us at the same problem we faced with the asking price method. The appraiser is typically hired by the realtor or seller, a merging of incentives cannot be rejected. Moreover, even though the appraiser could be independent, the realtor and seller could hire several appraisers, and picking the valuation which fits their strategy. Thus, the appraiser’s valuation in the data set could be the appraiser estimate which is closest to the asking price. It is worth noting however, that such strategy is costly, as they would have to pay for all the valuations not just the one that fits the asking price the best.

Thirdly, so called charm-pricing is not the only problem with using asking price and appraiser’s valuation as a baseline estimate for the value of an apartment. Even if the asking price is the true reservation price for the seller, a relative low ask still has a signal effect. This is a factor the appraisers’ valuation also shares, as it often is listed in a Norwegian online real estate advert together with the asking price. If apartments with a low appraiser’s valuation and asking price attracts more bidders, the coefficients from the asking price and appraiser’s valuation method will be over-estimated. This holds even if low valuation and low asking price is random and independent of incentives of the seller, realtor and appraiser.

Fourthly, although hedonic predicted prices should not suffer from the same issues as appraiser’s valuation and asking price, one has to be cautious in interpreting the number of bidders coefficient. If the hedonic model truly reflects a hedonic demand function of the attributes, and any residuals of the hedonic regression is due bidding round effects alone, the regression outputs indicates that the number of bidders raises the final selling price. However, it is easy to imagine how the number of bidders could correlate with unobserved variables. Even though the hedonic model is extensive, there might be a correlation between N and omitted variables in the hedonic model, which would bias the N-coefficient.

Finally, despite the potential issue with the hedonic addressed above, most weight should be put on this estimate of the N coefficient, this is because it is free from the signaling and incentive issues listed for the appraiser's valuation and asking price. Therefore, the size differences in the coefficient estimates might be interpreted in the direction of the existence of endogeneity in the asking price and appraiser's valuation spread methods.

Even with the most conservative interpretations of the number of bidders coefficients, as the effect in terms of deviation of price from the base values, the results are striking, and provides insights into the interaction between prices and number of bidders in real estate auctions.

6.1 Alternative approaches

To address the potential endogeneity issues with the appraiser's valuation and asking price methods. Approaches that specifically deals with endogeneity of estimators could be applied. One such approach is the method in the earlier mentioned paper of Genesove and Han(2011), of singling out the effect of the asking price on the final selling price, to further correct for it through maximum likelihood.

Another viable method would be the Instrumental-variable approach. This approach uses an additional variable called the instrument I, which is assumed to be uncorrelated with the problematic model error term, but is correlated with the endogenous variable (here number of bidders N) (Verbeek, 2012, p. 148). Through decomposing N into a problematic part and an unproblematic part, by predicting N by using I through OLS(first-stage), then regressing the spread on $\widehat{N}$(second-stage). $\widehat{N}$ in the second-stage regression is now assumed to be uncorrelated with the error term u_i. However, there are two important conditions for this method, these are: instrument relevance: $corr(I_i, N_i) \neq 0$ and instrument exogeneity: $corr(I_i, u_i) = 0$. (Stock and Watson, 2015, p. 472-473).

Finding a sufficiently strong instrument which is correlated with N should be relatively unproblematic. One approach could be to use the appraiser's valuation as an instrument for N in the asking price model. The first-stage regression (31): $N_i = \beta_0 + \beta_1 V_i + u_i$ where V is the appraiser's valuation and u_i is the error term, gives a F-statistic of 94, which is well above the rule of thumb for strong instruments of F>10 (Stock and Watson, 2015, p. 490). However the instrument exogeneity condition is unlikely to hold. If realtors and sellers use the appraiser's valuation as a guideline to set the asking price, then it is likely correlated with the error term in the second-stage regression.

In general, finding an instrument for N that satisfies the exogenity condition is challenging. One possible example of an exogenous instrument would be weather. The idea is that there would be more people attending viewings in sunny weather, than there would be on clouded or rainy days. Which would be likely to cause more bidders in the following auction. Weather should be uncorrelated with any apartment attributes, or the asking price as it is likely to have been determined before the weather forecast was available. Alas, any data on the date of the viewings is at the moment unavailable. Even with such data at hand, apartments are likely to have multiple viewing dates, thus making the impact of weather on number of bidders not straight forward.

6.2 Stochastic number of bidders

As an extension to the analysis in the previous sections that indicated that the number of bidders is associated with higher final selling price, the following question arises: What determines the number of bidders? The theoretical framework section provides a hypothesis: The number of bidders is stochastic.

To evaluate this hypothesis, I take advantage of apartments being sold more than once. To identify these, the cadaster numbers included in the data-set provide unique apartment specific identifiers. This method results in the identification of a total of 61 apartments that have been sold twice through the time-span of the data-set.

If the number of bidders is indeed stochastic, the number of bidders for the second sale will be uncorrelated with the number of bidders of the first sale. To test this the following regression model is employed: (32): $N_{h,2} = \beta_0 + \beta_1 N_{h,1} + u_h$. N_2 and N_1 are the number of bidders in sale 2 and 1for apartment h, u_h is the residual term with assumed $u_i \sim (0, \sigma^2)$. Note that omitting the asking price of sale 1 and 2, does not lead to omitted variable bias because the asking price of sale 1 is not a determinant of N_2 and the asking price of sale 2 should be uncorrelated with N_1 (Stock and Watson, 2015, p. 231). Moreover, note that the constant is common, this specification is not unit-specific.

Table 9: Regression results stochastic number of bidders

	$Numberofbidders_2$
$Numberofbidders_1$	0.3850^{***}
	(0.0980)
Constant	1.7367^{***}
	(0.3415)
Observations	61
R^2	0.14

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The hypothesis of number of bidder being stochastic amounts to the following:

$H_0\colon \beta_1 = 0 \; vs. H_1\colon \beta_1 \neq 0$ .Thus regression output leads us to reject H_0 at the 1% significance level. This assumes that 61 observations is sufficiently large for the Central Limit Theorem to hold, and the t-values of the regression to be approximately normally distributed.

This constitutes evidence for the number of bidders being significantly dependent on apartment specific attributes. It is unlikely that number of bidders between the sales has any causal link, since the number of bidders in a housing auction is not public knowledge. The regression result thus indicates that the arrival of bidders is not purely stochastic but correlate within apartments because certain types of apartments in Oslo attracts more bidders.

The findings of this section may be interpreted as indication of the existence of a deterministic component to N connected to the apartment itself. The implication for the model in the theoretical framework, is that the random portion of N, drawn from the Poisson distribution, is the portion over or under the apartment-specific N. In other words the arrival rate λ, is unit specific.

6.3 Apartment fixed effects

Given the results in 6.2, we now turn to a unit specific-effects set-up of the effect of the number of bidders on the final selling price.

Through the same method employed in 6.2, I identify apartments sold more than once.

By including intercepts for each apartment the method controls for omitted variables that vary across apartments but not across time (Stock and Watson, 2015, p. 410). This allows us to control for apartment specific attributes, which was not possible in the cross-sectional regressions. However, this process comes at a cost, as there are only 122 transactions for 61 apartments, it limits the degrees of freedom of the regression significantly. Moreover analyses on the appraiser's valuation spread and hedonic price spread cannot be performed, as there is a large amount of the 122 transactions that is missing either appraiser's valuation or variables included in the hedonic predictive model. Panel data for hedonic spread and appraisers valuation spread are unbalanced and the degrees of freedom are too limited. This leaves the asking price spread, which allows apartment fixed effects regression on a data set of 117 sales of 59 apartments. The regression model is the following:

(34): $S_{ht} = \beta_0 + \beta_1 N_{ht} + \gamma_1 H1_h + \gamma_2 H2_h + \cdots + \gamma_{j-1} Hj - 1_h + u_{ht}$

In which S_{ht} is the spread of apartment h in sale t,

where (34): $S_{ht} = \frac{Price_{ht} - V_{ht}}{Vht}$, $V_{ht} = askingprice_{ht}$ N_{ht} is the number of bidders of apartment h in transaction t, H1 through Hj-1 is apartment specific dummies, where j is the number of apartments, u is the error term. In case there are any correlation of u within sales of the same apartments, cluster robust standard error has been employed.

Table 10: Regression results, apartment fixed effects

	(1)
	Asking price spread
Number of Bidders	0.0180***
	(0.0048)
Observations	117
Groups	59
Within R^2	0.19
Between R^2	0.49
Overall R^2	0.39

Cluster robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From the regression output we observe that controlling for apartment fixed effects, one extra bidder is associated with 1.8 percentage point higher asking price spread. The coefficient is statistically significant even at the 1 percentage level. This finding further solidifies the results in section 5. However the coefficient is smaller than found in section 5.1, which indicates that might be unit specific effects not accounted for in 5.1. Moreover there might be the same apartment specific effects for appraiser's valuation spread and hedonic price spread. Due to the data restraints mentioned above we cannot apply methods to control for these. However if the apartment specific effects are similar across the methods, they are likely not to threaten the significance of the estimates.

7 Robustness check

To check whether the coefficients for the number of bidders on S reported in chapter 5.1, 5.2 and 5.3 are robust to the inclusion/exclusion of the observations identified in section 4.5. I have repeated the uncontrolled regressions when these observations are dropped. To repeat the reasoning in section 4.5, there are 373 observations, where a bidder currently having the highest bid and it being above the asking price, raises his the bid. The robustness check is done to address these observations as potential errors in input.

Table 11: Regression output robustness check.

	Identified observations not dropped			Identified observations dropped		
	S_a	S_h	S_p	S_a	S_v	S_p
Number of bidders	0.0246***	0.0235***	0.0067***	0.0250***	0.0239***	0.0068***
	(0.0005)	(0.0006)	(0.0008)	(0.0005)	(0.0006)	(0.0008)
Constant	-0.0005	-0.0122***	-0.0175***	-0.0023	-0.0141***	-0.0178***
	(0.0016)	(0.0020)	(0.0033)	(0.0017)	(0.0022)	(0.0036)
Observations	4509	3293	4509	4136	3018	4136
Adjusted R^2	0.47	0.38	0.01	0.47	0.38	0.01

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As earlier, S_a is the asking price spread, S_v is the appraiser's valuation spread, and S_p is hedonic predicted prices spread.

The regression output indicates that the number of bidders coefficients in the main regressions are in large part robust to the exclusion/inclusion of the identified observations.

8 Conclusion

In conclusion, this thesis finds that a higher number of bidders is associated with a higher final selling price, when measured as the spread from the asking price, appraiser's valuation and the hedonic predicted price. The coefficients of the number of bidders on these spreads are statistically significant at 1% level for all three methods. The estimate of the coefficient on the asking price spread and the appraiser's valuation spread are similar, however the hedonic spread coefficient is substantially lower. The latter estimate should bare the most weight as the other two might suffer from issues of endogeneity. Furthermore, the assumption of number of bidders being stochastic put forward in the theoretical framework, is rejected at the 1% significance level, when regressing the number of bidders between sales for the same apartment on each other. This leads to an inquiry of apartment specific effects, the thesis finds that a higher number of bidders is associated with higher asking price spread while controlling for apartment specific effects, however the coefficient is lower than in the original set-up. Due to limitations in the data set, unfortunately the same method cannot be performed on the two other spreads.

Even when being careful in interpreting the number of bidders coefficients as evidence of effect on the final selling price alone, the findings should provide interesting insights to the English auction for housing markets, and the effect number of bidders has.

References

Albrecht J., P.A. Gautier and S. Vroman (2015): Directed Search in the Housing Market, *Review of Economic Dynamics 19, 218-231*

Albrecht J., A. Anderson, E. Smith and S. Vroman (2007): Opportunistic Matching in the Housing Market, *International Economic Review* 48(2), 641-664

Ashenfelter O. and D. Genesove(1992): Testing for Price Anomalies in Real Estate Auctions, *National Bureau of Economic Research, Inc. Working Paper* No. 4036

Box, G. E. P., and D. R. Cox. (1964): An analysis of transformations, *Journal of the Royal Statistical Society Series B(Methodological)* 26(2), 211-252

Can A. (1992): Specification and Estimation of Hedonic Housing Price Models, *Regional Science and Urban Economics* 22, 453-474

Carrillo P.E (2013): To Sell or Not to Sell: Measuring the Heat of the Housing Market, *Real Estate Economics* 41(2), 310-346

Cropper M.L, L.B Deck and K.E McConnell (1988): On the Choice of Functional Form for Hedonic Price Functions, *The Review of Economics and Statistics* 70 (4), 668-675

Furnham A. and H.C Boo (2011): A Literature Review of the Anchoring Effect, *The Journal of Socio-Economics* 40(1), 35-42

Genesove D. and L. Han (2011) Measuring the Thinness of Real Estate Markets, unpublished manuscript[3], Hebrew University of Jerusalem, CEPR and Rotman School of Management University of Toronto

[3] Downloaded from https://iweb.cerge-ei.cz/pdf/events/papers/111003_t.pdf at 18.04.2017

Genesove D. and L. Han (2012): Search and Matching in the Housing Market, *Journal of Urban Economics 72* , 31-45

Han, L. and W.C. Strange (2016): What is the Role of the Asking Price for a house, *Journal of Urban Economics 93* , 115-130.

Han, L. and W.C. Strange (2015): The Microstructure of Housing Markets: Search, Bargaining, and Brokerage. In G. Duranton, J.V Henderson and W.C Strange(eds.) *Handbook of Regional and Urban Economics Volume 5.* Amsterdam: North Holland

Han, L. and W.C. Strange (2014): Bidding Wars for Houses, *Real Estate Economics 42(1)*, 1-32

Haurin D.R., J.L. Haurin and T. Naudauld (2010): List Prices, Sale Prices and Marketing Time: An Application to U.S Housing Markets, *Real Estate Economics* 38(4), 659-685

Holt, C. A. Jr.. (1979). A Theory of Signaling Auctions, *Discussion Paper No. 110*, Center for Economic Research, Department of Economics, University of Minnesota[4]

Hoover K.D and S.J. Perez (1999): Data mining reconsidered: Encompassing and the General-to-Specific Approach to Specification Search, *Econometrics Journal* 2, 167–191

Jensen A. (2016): The Effect of Commuting Time on House Prices, Master Thesis, Department of Economics University of Oslo

Knight J.R, C.F Sirmans and G.F Turnbull (1994): List Price Signaling and Buyer Behavior in the Housing Market, *Journal of Real Estate Finance and Economics 9*, 177-192

Krainer J. (2001): A Theory of Liquidity in Residential Real Estate Markets, *Journal of Urban Economics* 49, 32-53

McAfee R.P, D.C Quan and D.R Vincent (2002): How to Set Minimum Acceptable Bids, With an Application to Real Estate Auctions, *The Journal of Industrial Economics* L(4), 391-416

[4] Downloaded from http://hdl.handle.net/11299/54994 at 18.04.2017

McAfee, R.P and J. McMillian (1987a): Auctions and Bidding, *Journal of Economic Literature* 25 (2), 699-738

McAfee, R.P and J. McMillian (1987b): Auctions with a Stochastic Number of Bidders, *Journal of Economic Theory* 43, 1-19

Nenov P.T, E.R Larsen and D.E Sommervoll (2016): Thick-market Effects, Housing Heterogeneity and the Determinants of Transaction Seasonality, *The Economic Journal* 126(598), 2403-2423

Northcraft G.B and M.A Neal (1987): Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions, *Organizational Behavior and Human Decision Processes 39*, 84-97

Rosen S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy* 82(1), 34-55

Sheppard S. (1999): Hedonic Analysis of Housing Markets. In E.S Mills P. Cheshire(eds.) *Handbook of Regional and Urban Economics Volume 3,* Amsterdam: North Holland

Sommervoll D.E, E.F Berg and A.R Skyvulstad(2015): Slumbering Cobra, Dead or Deadly?: Jump Bids in Real Estate Auctions, unpublished manuscript, Norwegian University of Life Sciences

Skjærholt A. (2015) Behavioral Biases in Housing Markets - The Case of Oslo, master thesis, Department of Economics University of Oslo

Stock J.H and M.K Watson (2015) *Introduction to Econometrics*, updated third edition, Harlow Essex: Pearson Education Limited

Theisen T. (2016) Efficiency and Justice in the Market for Cooperative Dwellings, *International Real Estate Review* 19(3), 297-326

Thompson B. (1995): Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply Here: A Guidelines Editorial, *Educational and Psychological Measurement* 55(4), 525-534

Verbeek M. (2012): *A Guide to Modern Econometrics*, fourth edition Chichester West Sussex: John Wiley & Sons Ltd.

Yinger J. (1981): A Search Model of Real Estate Broker Behavior, *The American Economic Review* 71(4), 591-605

Additional references:

NEF (2016): *Regler for budgivning(Rules for bidding rounds)*(Online) Available at: http://www.nef.no/forbruker/regler-for-budgivning/ (Accessed: 22.04.2017)

Eiendom Norge (2017): *Eiendom Norges boligprisstatistikk April 2017(The Norwegian Housing Price Statistics April 2017)*

Wikipedia (2017): *English auction (Online)* Available at: https://en.wikipedia.org/wiki/English_auction (Accessed: 09.05.2017)

Appendix

A.1 Regression output from asking price and appraiser's valuation model

Table 12: Regression output for section 5.1 asking price spread

	(1) S_a	(2) S_a	(3) S_a	(4) S_a
Number of bidders	0.0246***	0.0238***	0.0233***	0.0221***
	(0.0005)	(0.0005)	(0.0005)	(0.0004)
Size		-0.0005***	-0.0004***	-0.0004***
		(0.0000)	(0.0000)	(0.0000)
ALNA			0.0320*	0.0223
			(0.0141)	(0.0141)
BJERKE			0.0218	0.0152
			(0.0138)	(0.0138)
FROGNER			0.0008	-0.0083
			(0.0137)	(0.0137)
GAMLEOSLO			0.0316*	0.0188
			(0.0132)	(0.0133)
GRORUD			0.0370*	0.0299*
			(0.0145)	(0.0144)
GRUNERLØKKA			0.0242	0.0127
			(0.0133)	(0.0133)
MARKA			(.)	(.)
			(.)	(.)
NORDREAKER			0.0277*	0.0183
			(0.0140)	(0.0140)
NORDSTRAND			0.0218	0.0112
			(0.0140)	(0.0140)
SAGENE			0.0270*	0.0162
			(0.0137)	(0.0137)
SENTRUM			0.0076	-0.0006
			(0.0216)	(0.0213)
STHANSHAUGEN			0.0174	0.0099
			(0.0139)	(0.0139)
STOVNER			0.0295*	0.0230
			(0.0147)	(0.0145)
SØNDRENORDSTRAND			0.0160	0.0081
			(0.0166)	(0.0167)
ULLERN			0.0062	0.0037
			(0.0140)	(0.0140)
VESTREAKER			0.0108	0.0063
			(0.0139)	(0.0139)
ØSTENSJØ			0.0401**	0.0310*
			(0.0140)	(0.0140)
year=2013				(.)
				(.)
year=2014				0.0089*
				(0.0043)
year=2015				0.0342***
				(0.0042)
year=2016				0.0532***
				(0.0045)
year=2017				-0.0166
				(0.0086)
month=1				(.)
				(.)
month=2				0.0026
				(0.0044)
month=3				-0.0002

				(0.0042)
month=4				0.0110*
				(0.0043)
month=5				0.0116**
				(0.0043)
month=6				0.0049
				(0.0042)
month=7				0.0001
				(0.0070)
month=8				0.0173***
				(0.0044)
month=9				0.0150***
				(0.0045)
month=10				0.0051
				(0.0044)
month=11				-0.0013
				(0.0043)
month=12				-0.0046
				(0.0048)
Constant	-0.0005	0.0367***	0.0101	-0.0157
	(0.0016)	(0.0035)	(0.0139)	(0.0147)
Observations	4509	4509	4509	4509
Adjusted R^2	0.47	0.48	0.50	0.55

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (.) omitted due to collinearity

Table 13: Regression output for section 5.2 appraiser's valuation spread

	(1)	(2)	(3)	(4)
	S_v	S_v	S_v	S_v
Number of bidders	0.0235***	0.0228***	0.0222***	0.0212***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Size		-0.0005***	-0.0004***	-0.0004***
		(0.0001)	(0.0001)	(0.0001)
ALNA			0.0363	0.0281
			(0.0205)	(0.0201)
BJERKE			0.0270	0.0225
			(0.0201)	(0.0198)
FROGNER			0.0054	-0.0039
			(0.0201)	(0.0198)
GAMLEOSLO			0.0394*	0.0252
			(0.0196)	(0.0193)
GRORUD			0.0455*	0.0389
			(0.0210)	(0.0205)
GRUNERLØKKA			0.0267	0.0160
			(0.0196)	(0.0193)
MARKA			(.)	(.)
			(.)	(.)
NORDREAKER			0.0329	0.0255
			(0.0204)	(0.0200)
NORDSTRAND			0.0291	0.0188
			(0.0204)	(0.0201)
SAGENE			0.0302	0.0202
			(0.0201)	(0.0198)
SENTRUM			0.0120	0.0092
			(0.0333)	(0.0308)
STHANSHAUGEN			0.0210	0.0126
			(0.0203)	(0.0200)
STOVNER			0.0331	0.0264
			(0.0207)	(0.0203)
SØNDRENORDSTRAND			0.0267	0.0159
			(0.0234)	(0.0233)
ULLERN			0.0056	0.0033
			(0.0204)	(0.0201)
VESTREAKER			0.0125	0.0048
			(0.0203)	(0.0200)
ØSTENSJØ			0.0441*	0.0355
			(0.0204)	(0.0201)
year=2013				(.)
				(.)
year=2014				0.0203**

				(0.0062)
year=2015				0.0512***
				(0.0061)
year=2016				0.0697***
				(0.0071)
month=1				(.)
				(.)
month=2				0.0085
				(0.0050)
month=3				0.0052
				(0.0048)
month=4				0.0172***
				(0.0050)
month=5				0.0089
				(0.0052)
month=6				0.0120*
				(0.0051)
month=7				0.0077
				(0.0089)
month=8				0.0144**
				(0.0052)
month=9				0.0190***
				(0.0054)
month=10				0.0053
				(0.0055)
month=11				0.0086
				(0.0053)
month=12				0.0033
				(0.0061)
Constant	-0.0122***	0.0224***	-0.0083	-0.0490*
	(0.0020)	(0.0045)	(0.0203)	(0.0212)
Observations	3293	3293	3293	3293
Adjusted R^2	0.38	0.40	0.41	0.47

Robust standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (.) omitted due to collinearity

A.2 Selection of functional form and regressors in the hedonic model

Table 14: Functional form specification

	(1) Price	(2) Price	(3) Price	(4) Price
commondebt	-0.168***	-0.169***	-0.170***	-0.170***
	(-4.9269)	(-4.9380)	(-4.9666)	(-4.9666)
numbedroom	431722.0	536310.5***	534025.2***	534025.2***
	(0.9100)	(4.2277)	(4.2056)	(4.2056)
numbedroom2	-40696.1	-91378.9**	-91120.2**	-91120.2**
	(-0.1650)	(-2.7170)	(-2.7082)	(-2.7082)
numbedroom3	-7412.5			
	(-0.1927)			
floor	43159.8*	33131.4**	47309.2***	47309.2***
	(2.0003)	(2.6580)	(7.2138)	(7.2138)
floor2	-991.9	1626.1		
	(-0.1943)	(1.1327)		
floor3	164.6			
	(0.5458)			
size	25391.3	26771.2**	26913.8**	26913.8**
	(1.4524)	(3.0282)	(3.0328)	(3.0328)
Size2	153.7	135.7**	135.5**	135.5**
	(0.6331)	(2.6508)	(2.6461)	(2.6461)

Size3	-0.0663			
	(-0.0625)			
buildyear	17483637.7***	17490624.2***	17405616.5***	17405616.5***
	(7.4690)	(7.4988)	(7.4581)	(7.4581)
buildyear2	-9079.5***	-9083.2***	-9039.7***	-9039.7***
	(-7.5374)	(-7.5659)	(-7.5257)	(-7.5257)
buildyear3	1.571***	1.572***	1.565***	1.565***
	(7.6065)	(7.6342)	(7.5943)	(7.5943)
buildyear4				(.)
				(.)
year=2014	190568.5***	190168.3***	190299.9***	190299.9***
	(3.9574)	(3.9500)	(3.9474)	(3.9474)
year=2015	568703.3***	568163.5***	569015.3***	569015.3***
	(11.9581)	(11.9715)	(11.9716)	(11.9716)
year=2016	1124062.6***	1124082.0***	1125442.8***	1125442.8***
	(22.4780)	(22.4948)	(22.4725)	(22.4725)
year=2017	1919103.6***	1917970.1***	1915466.5***	1915466.5***
	(24.1333)	(23.7913)	(23.7496)	(23.7496)
month=1	(.)	(.)	(.)	(.)
	(.)	(.)	(.)	(.)
month=2	79380.4	79788.6	79919.1	79919.1
	(1.5741)	(1.5387)	(1.5423)	(1.5423)
month=3	56760.6	58064.7	59076.1	59076.1
	(1.4725)	(1.5045)	(1.5330)	(1.5330)
month=4	150788.4**	151364.6**	152386.1**	152386.1**
	(3.0351)	(3.0521)	(3.0789)	(3.0789)
month=5	146488.3***	146841.9***	148212.9***	148212.9***
	(3.6390)	(3.6478)	(3.6890)	(3.6890)
month=6	201551.3***	202303.4***	202383.7***	202383.7***
	(5.1209)	(5.1228)	(5.1276)	(5.1276)
month=7	280520.8***	280868.2***	282009.1***	282009.1***
	(4.7100)	(4.6971)	(4.7179)	(4.7179)
month=8	290205.1***	290672.4***	290937.9***	290937.9***
	(7.3810)	(7.3558)	(7.3676)	(7.3676)
month=9	302673.6***	303517.1***	303949.9***	303949.9***
	(6.7123)	(6.7469)	(6.7516)	(6.7516)
month=10	334451.7***	334329.2***	334331.6***	334331.6***
	(7.5321)	(7.5115)	(7.5173)	(7.5173)
month=11	385655.4***	386557.2***	386816.4***	386816.4***
	(8.9702)	(8.9453)	(8.9523)	(8.9523)
month=12	343311.4***	343844.2***	346211.4***	346211.4***
	(7.1807)	(7.1442)	(7.1985)	(7.1985)
garage	160443.2***	159937.2***	158088.3***	158088.3***
	(6.2973)	(6.5440)	(6.4702)	(6.4702)
coop	-7992.6	-7984.5	-7284.4	-7284.4
	(-0.3969)	(-0.4050)	(-0.3693)	(-0.3693)
balcony	-42033.4	-41814.8*	-45309.6*	-45309.6*
	(-1.7343)	(-1.9965)	(-2.1760)	(-2.1760)
aview	124524.1***	123025.9***	124970.2***	124970.2***
	(4.0959)	(4.1123)	(4.1883)	(4.1883)
fireplace	155925.5***	156429.0***	157310.8***	157310.8***
	(4.0160)	(4.0327)	(4.0556)	(4.0556)
needsrenovating	-259011.8***	-259808.6***	-259903.9***	-259903.9***
	(-7.2592)	(-7.1095)	(-7.1277)	(-7.1277)
renovated	130772.5**	130663.8**	130689.5**	130689.5**
	(3.0888)	(3.0917)	(3.0925)	(3.0925)
sunny	35945.1	35481.2	35426.1	35426.1
	(1.7498)	(1.7147)	(1.7098)	(1.7098)
heatingrating	-7942.2	-7814.6	-8040.9	-8040.9
	(-1.3069)	(-1.2682)	(-1.3065)	(-1.3065)
energyrating	-14642.7	-15003.6	-14660.4	-14660.4
	(-1.6518)	(-1.6799)	(-1.6385)	(-1.6385)
ALNA	1421528.8***	1416192.4***	1436868.9***	1436868.9***
	(4.2757)	(4.2365)	(4.2672)	(4.2672)
BJERKE	172122.1	169201.3	178150.0	178150.0
	(0.5152)	(0.5021)	(0.5242)	(0.5242)
FROGNER	716021.4*	716233.9*	724944.1*	724944.1*
	(2.1540)	(2.1149)	(2.1227)	(2.1227)
GAMLEOSLO	566720.3	567578.5	576802.4	576802.4
	(1.9155)	(1.8881)	(1.8990)	(1.8990)
GRORUD	1721856.6***	1714473.3***	1715368.8***	1715368.8***
	(4.5152)	(4.4658)	(4.4533)	(4.4533)
GRUNERLØKKA	655451.2*	655567.1*	659320.8*	659320.8*
	(2.1843)	(2.1540)	(2.1447)	(2.1447)
MARKA	(.)	(.)	(.)	(.)

	(.)	(.)	(.)	(.)
NORDREAKER	263679.1	258100.5	264963.1	264963.1
	(0.7734)	(0.7463)	(0.7605)	(0.7605)
NORDSTRAND	126877.4	118624.5	124335.7	124335.7
	(0.3560)	(0.3199)	(0.3334)	(0.3334)
SAGENE	729558.8*	727843.9*	734349.8*	734349.8*
	(2.2346)	(2.2080)	(2.2089)	(2.2089)
SENTRUM	-366253.6	-353762.3	-346383.7	-346383.7
	(-0.5242)	(-0.4978)	(-0.4879)	(-0.4879)
STHANSHAUGEN	876412.9**	874303.4**	883021.3**	883021.3**
	(2.6419)	(2.6058)	(2.6100)	(2.6100)
STOVNER	1408127.6***	1402426.2***	1403287.5***	1403287.5***
	(4.2174)	(4.2047)	(4.1687)	(4.1687)
SØNDRENORDSTRAND	1326370.1***	1320292.6***	1331382.0***	1331382.0***
	(3.5805)	(3.5656)	(3.5666)	(3.5666)
ULLERN	1425819.2***	1423623.3***	1427864.4***	1427864.4***
	(3.7932)	(3.7202)	(3.7119)	(3.7119)
VESTREAKER	1024381.2*	1033185.5*	1040003.1*	1040003.1*
	(2.2940)	(2.3761)	(2.3824)	(2.3824)
ØSTENSJØ	608008.8	599897.2	606176.1	606176.1
	(1.7795)	(1.7447)	(1.7503)	(1.7503)
ALNAxliving	-34397.0***	-34304.1***	-34491.3***	-34491.3***
	(-6.0379)	(-5.9665)	(-5.9614)	(-5.9614)
BJERKExliving	-7113.0	-7053.9	-7129.5	-7129.5
	(-1.2288)	(-1.2067)	(-1.2119)	(-1.2119)
FROGNERxliving	3903.2	3897.8	3804.8	3804.8
	(0.6731)	(0.6621)	(0.6423)	(0.6423)
GAMLEOSLOxliving	-8804.2	-8820.7	-8921.9	-8921.9
	(-1.7120)	(-1.6844)	(-1.6904)	(-1.6904)
GRORUDxliving	-40234.0***	-40123.3***	-40030.5***	-40030.5***
	(-6.4683)	(-6.3826)	(-6.3542)	(-6.3542)
GRUNERLØKKAxliving	-8310.2	-8323.8	-8360.4	-8360.4
	(-1.5932)	(-1.5714)	(-1.5665)	(-1.5665)
MARKAxliving	(.)	(.)	(.)	(.)
	(.)	(.)	(.)	(.)
NORDREAKERxliving	1571.3	1660.5	1614.8	1614.8
	(0.2670)	(0.2786)	(0.2694)	(0.2694)
NORDSTRANDxliving	-2598.9	-2483.7	-2507.9	-2507.9
	(-0.4295)	(-0.3980)	(-0.3999)	(-0.3999)
SAGENExliving	-5329.7	-5293.2	-5370.5	-5370.5
	(-0.9286)	(-0.9145)	(-0.9222)	(-0.9222)
SENTRUMxliving	26354.1	26072.6	25869.8	25869.8
	(1.4745)	(1.4375)	(1.4305)	(1.4305)
STHANSHAUGENxliving	-3962.2	-3935.4	-4040.6	-4040.6
	(-0.6837)	(-0.6717)	(-0.6853)	(-0.6853)
STOVNERxliving	-37379.9***	-37295.5***	-37234.7***	-37234.7***
	(-6.6199)	(-6.5725)	(-6.5160)	(-6.5160)
SØNDRENORDSTRANDxliving	-38403.7***	-38281.9***	-38438.4***	-38438.4***
	(-6.3396)	(-6.2794)	(-6.2654)	(-6.2654)
ULLERNxliving	-15877.3*	-15812.3*	-15841.2*	-15841.2*
	(-2.5457)	(-2.5061)	(-2.4991)	(-2.4991)
VESTREAKERxliving	-13066.6	-13183.5	-13190.4	-13190.4
	(-1.8784)	(-1.9308)	(-1.9242)	(-1.9242)
ØSTENSJØxliving	-16326.4**	-16199.9**	-16262.4**	-16262.4**
	(-2.7829)	(-2.7388)	(-2.7340)	(-2.7340)
Constant	-1.12197e+10***	-1.12242e+10***	-1.11689e+10***	-1.11689e+10***
	(-7.4021)	(-7.4328)	(-7.3918)	(-7.3918)
Observations	4509	4509	4509	4509
R^2	0.83	0.83	0.83	0.83

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (.) omitted due to collinearity

In model 1: Second and third power are included of numberofbedrooms, size, buildyear and floor.

In model 2: Third power are dropped of numberofbedrooms, size and floor

In model 3: Second power of floor is dropped

In model 4: Fourth power of buildyear is included, however it is omitted by Stata because of collinearity.

Table 15: Selecting regressors from proxy group.

	(1) Price	(2) Price	(3) Price	(4) Price	(5) Price
commondebt	-0.170***	-0.169***	-0.171***	-0.171***	-0.172***
	(-4.9666)	(-4.9650)	(-5.0325)	(-5.0108)	(-5.0036)
numbedroom	534025.2***	536406.5***	536301.0***	534251.7***	527099.2***
	(4.2056)	(4.2130)	(4.2173)	(4.2003)	(4.1498)
numbedroom2	-91120.2**	-91510.9**	-91388.5**	-90885.2**	-89756.3**
	(-2.7082)	(-2.7143)	(-2.7142)	(-2.6988)	(-2.6646)
floor	47309.2***	48119.6***	48280.4***	48636.9***	47545.8***
	(7.2138)	(7.2934)	(7.3275)	(7.4326)	(7.3859)
size	26913.8**	26892.0**	27161.4**	27551.9**	27096.9**
	(3.0328)	(3.0288)	(3.0418)	(3.0577)	(3.0030)
size2	135.5**	135.3**	134.7**	135.1**	136.1**
	(2.6461)	(2.6428)	(2.6331)	(2.6431)	(2.6530)
buildyear	17405616.5***	17869815.8***	17789567.1***	18323481.1***	18421947.1***
	(7.4581)	(7.7083)	(7.6945)	(7.9363)	(8.0305)
buildyear2	-9039.7***	-9277.9***	-9236.2***	-9515.8***	-9568.2***
	(-7.5257)	(-7.7754)	(-7.7615)	(-8.0097)	(-8.1069)
buildyear3	1.565***	1.605***	1.598***	1.647***	1.656***
	(7.5943)	(7.8439)	(7.8292)	(8.0843)	(8.1852)
year=2013	(.)	(.)	(.)	(.)	(.)
	(.)	(.)	(.)	(.)	(.)
year=2014	190299.9***	191435.2***	192441.0***	193338.3***	188196.7***
	(3.9474)	(3.9719)	(3.9919)	(4.0073)	(3.9076)
year=2015	569015.3***	570745.9***	572931.0***	573559.8***	565797.0***
	(11.9716)	(12.0136)	(12.0491)	(12.0551)	(11.9524)
year=2016	1125442.8***	1128907.7***	1129989.1***	1131166.9***	1123229.1***
	(22.4725)	(22.5669)	(22.5859)	(22.6181)	(22.5670)
year=2017	1915466.5***	1912873.6***	1909763.4***	1904345.7***	1881874.1***
	(23.7496)	(23.7584)	(23.7345)	(23.6796)	(23.8793)
month=1	(.)	(.)	(.)	(.)	(.)
	(.)	(.)	(.)	(.)	(.)
month=2	79919.1	78986.3	77938.8	76829.6	76794.3
	(1.5423)	(1.5252)	(1.5061)	(1.4854)	(1.4827)
month=3	59076.1	58436.0	57292.7	57496.3	58803.4
	(1.5330)	(1.5175)	(1.4896)	(1.4946)	(1.5269)
month=4	152386.1**	151124.5**	150580.4**	152112.6**	151083.5**
	(3.0789)	(3.0560)	(3.0465)	(3.0808)	(3.0570)
month=5	148212.9***	147830.0***	148584.7***	149787.4***	149662.9***
	(3.6890)	(3.6795)	(3.7037)	(3.7292)	(3.7258)
month=6	202383.7***	201827.9***	201705.3***	203040.6***	202020.9***
	(5.1276)	(5.1153)	(5.1203)	(5.1561)	(5.1269)
month=7	282009.1***	282986.6***	286519.5***	288594.8***	286375.6***
	(4.7179)	(4.7293)	(4.8070)	(4.8541)	(4.8082)
month=8	290937.9***	291483.5***	292164.6***	293368.3***	292975.4***
	(7.3676)	(7.3861)	(7.4156)	(7.4363)	(7.4178)
month=9	303949.9***	304787.5***	304936.2***	304782.6***	305183.1***
	(6.7516)	(6.7689)	(6.7765)	(6.7797)	(6.7674)
month=10	334331.6***	334397.6***	333337.0***	333523.8***	334415.5***
	(7.5173)	(7.5115)	(7.4895)	(7.4902)	(7.4941)

month=11	386816.4***	386891.5***	385807.3***	386390.9***	385868.1***
	(8.9523)	(8.9631)	(8.9407)	(8.9578)	(8.9380)
month=12	346211.4***	346752.4***	343940.6***	344590.8***	343219.7***
	(7.1985)	(7.2227)	(7.1797)	(7.1935)	(7.1757)
garage	158088.3***	154568.6***	155671.1***	157596.1***	159644.5***
	(6.4702)	(6.4579)	(6.5088)	(6.5951)	(6.6532)
coop	-7284.4	-6395.4	-4240.1	-3280.2	-5570.3
	(-0.3693)	(-0.3227)	(-0.2145)	(-0.1663)	(-0.2841)
balcony	-45309.6*	-44197.9*	-38196.0	-36805.4	
	(-2.1760)	(-2.1348)	(-1.8690)	(-1.7997)	
aview	124970.2***	125097.9***	125161.6***	124886.3***	125454.2***
	(4.1883)	(4.1895)	(4.1939)	(4.1913)	(4.2151)
fireplace	157310.8***	157657.0***	158039.4***	157556.6***	152022.9***
	(4.0556)	(4.0691)	(4.0802)	(4.0696)	(3.9488)
needsrenovating	-259903.9***	-262079.3***	-266440.4***	-268600.6***	-269080.8***
	(-7.1277)	(-7.1932)	(-7.3374)	(-7.4141)	(-7.4177)
renovated	130689.5**	130620.4**	128849.5**	127724.5**	126329.9**
	(3.0925)	(3.0888)	(3.0447)	(3.0151)	(2.9941)
sunny	35426.1	34838.0			
	(1.7098)	(1.6849)			
heatingrating	-8040.9				
	(-1.3065)				
energyrating	-14660.4	-15136.2	-15654.8		
	(-1.6385)	(-1.6969)	(-1.7528)		
ALNA	1436868.9***	1424307.6***	1422936.2***	1468722.7***	1428916.4***
	(4.2672)	(4.2121)	(4.1239)	(4.2188)	(4.1207)
BJERKE	178150.0	184444.7	192320.4	231770.5	193968.9
	(0.5242)	(0.5411)	(0.5536)	(0.6606)	(0.5548)
FROGNER	724944.1*	708915.9*	708929.0*	751036.7*	727830.6*
	(2.1227)	(2.0682)	(2.0294)	(2.1243)	(2.0655)
GAMLEOSLO	576802.4	565284.5	562785.9	609433.4	576319.6
	(1.8990)	(1.8543)	(1.8024)	(1.9287)	(1.8321)
GRORUD	1715368.8***	1706812.7***	1714801.2***	1754950.9***	1714732.2***
	(4.4533)	(4.4193)	(4.3709)	(4.4354)	(4.3796)
GRUNERLØKKA	659320.8*	648439.8*	648087.6*	686764.9*	662598.4*
	(2.1447)	(2.1027)	(2.0525)	(2.1473)	(2.0827)
MARKA	(.)	(.)	(.)	(.)	(.)
	(.)	(.)	(.)	(.)	(.)
NORDREAKER	264963.1	251371.8	254474.0	292135.2	254812.7
	(0.7605)	(0.7203)	(0.7161)	(0.8144)	(0.7137)
NORDSTRAND	124335.7	107460.0	108562.4	156412.8	116816.7
	(0.3334)	(0.2884)	(0.2867)	(0.4108)	(0.3085)
SAGENE	734349.8*	719752.8*	723701.2*	770968.7*	742263.9*
	(2.2089)	(2.1586)	(2.1277)	(2.2469)	(2.1710)
SENTRUM	-346383.7	-398460.7	-394496.5	-320254.8	-320956.4
	(-0.4879)	(-0.5507)	(-0.5426)	(-0.4310)	(-0.4355)
STHANSHAUGEN	883021.3**	870745.9*	871343.6*	910423.3**	881537.1*
	(2.6100)	(2.5671)	(2.5193)	(2.6035)	(2.5333)
STOVNER	1403287.5***	1388548.3***	1406415.6***	1450060.6***	1411403.3***
	(4.1687)	(4.1174)	(4.0933)	(4.1803)	(4.0885)
SØNDRENORDSTRAND	1331382.0***	1319082.6***	1320909.8***	1357493.4***	1337299.6***
	(3.5666)	(3.5333)	(3.4940)	(3.5604)	(3.5119)
ULLERN	1427864.4***	1416357.7***	1416349.5***	1465075.5***	1437064.5***
	(3.7119)	(3.6794)	(3.6251)	(3.7275)	(3.6723)
VESTREAKER	1040003.1*	1019949.2*	1021793.8*	1065714.5*	1042458.9*
	(2.3824)	(2.3408)	(2.3205)	(2.4112)	(2.3700)
ØSTENSJØ	606176.1	596120.4	610545.5	649417.4	614119.9
	(1.7503)	(1.7168)	(1.7277)	(1.8178)	(1.7257)
ALNAxliving	-34491.3***	-34507.1***	-34638.3***	-35151.4***	-34673.5***
	(-5.9614)	(-5.9622)	(-5.8735)	(-5.8564)	(-5.8042)
BJERKExliving	-7129.5	-7317.8	-7483.7	-7982.5	-7529.8
	(-1.2119)	(-1.2441)	(-1.2500)	(-1.3101)	(-1.2405)
FROGNERxliving	3804.8	3885.9	3736.9	3243.0	3498.9
	(0.6423)	(0.6558)	(0.6197)	(0.5276)	(0.5714)
GAMLEOSLOxliving	-8921.9	-8921.7	-9006.7	-9594.9	-9220.0
	(-1.6904)	(-1.6909)	(-1.6700)	(-1.7416)	(-1.6822)
GRORUDxliving	-40030.5***	-40052.3***	-40308.7***	-40784.4***	-40327.5***
	(-6.3542)	(-6.3572)	(-6.2993)	(-6.2723)	(-6.2564)
GRUNERLØKKAxliving	-8360.4	-8325.6	-8445.4	-8895.7	-8654.1
	(-1.5665)	(-1.5602)	(-1.5485)	(-1.5956)	(-1.5609)
MARKAxliving	(.)	(.)	(.)	(.)	(.)
	(.)	(.)	(.)	(.)	(.)
NORDREAKERxliving	1614.8	1594.4	1405.3	973.7	1460.3
	(0.2694)	(0.2661)	(0.2306)	(0.1571)	(0.2367)
NORDSTRANDxliving	-2507.9	-2478.8	-2581.7	-3162.1	-2617.5

	(-0.3999)	(-0.3955)	(-0.4054)	(-0.4897)	(-0.4077)
SAGENExliving	-5370.5	-5337.9	-5531.6	-6064.2	-5720.5
	(-0.9222)	(-0.9165)	(-0.9332)	(-1.0054)	(-0.9521)
SENTRUMxliving	25869.8	27489.7	27289.2	26573.8	26396.3
	(1.4305)	(1.4868)	(1.4730)	(1.4034)	(1.4068)
STHANSHAUGENxliving	-4040.6	-4004.6	-4161.4	-4603.8	-4275.9
	(-0.6853)	(-0.6792)	(-0.6935)	(-0.7533)	(-0.7030)
STOVNERxliving	-37234.7***	-37215.1***	-37513.1***	-38073.4***	-37590.0***
	(-6.5160)	(-6.5178)	(-6.4523)	(-6.4323)	(-6.3876)
SØNDRENORDSTRAND xliving	-38438.4***	-38289.3***	-38417.1***	-38881.0***	-38515.9***
	(-6.2654)	(-6.2524)	(-6.1857)	(-6.1597)	(-6.1196)
ULLERNxliving	-15841.2*	-15813.5*	-15942.2*	-16541.2*	-16106.2*
	(-2.4991)	(-2.4962)	(-2.4784)	(-2.5356)	(-2.4817)
VESTREAKERxliving	-13190.4	-13136.7	-13280.9	-13815.0*	-13434.3
	(-1.9242)	(-1.9181)	(-1.9156)	(-1.9694)	(-1.9269)
ØSTENSJØxliving	-16262.4**	-16263.2**	-16497.5**	-16939.4**	-16383.4**
	(-2.7340)	(-2.7343)	(-2.7277)	(-2.7498)	(-2.6716)
Constant	-1.11689e+10***	-1.14704e+10***	-1.14190e+10***	-1.17589e+10***	-1.18205e+10***
	(-7.3918)	(-7.6421)	(-7.6288)	(-7.8639)	(-7.9551)
Observations	4509	4509	4509	4509	4509
R^2	0.83	0.83	0.83	0.83	0.83

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (.) omitted due to collinearity

In model 1: Unrestricted with the final functional form

In model 2: Heatingrating is dropped with t-value=-1.31

In model 3: Sunny is dropped with t -value=1.68

In model 4: Energyrating is dropped with t-value:-1.75

In model 5: Balcony is dropped with t-value:-1.79

Model 5 is used to predict the prices.

A.3 Test for overfitting when using zip code as location dummy

Table 16: Measures of fit for hedonic model in predicting transaction prices, with zip-code as location variable.

	R^2	Hit-rate*
In sample	0.93	96.7 %
Out of sample	0.81	88.5 %

*hit-rate(share of sales prices within 20% of predicted prices)

Sample split is conducted using a random number generator, splitting the data 70%(in sample) and 30%(out of sample). The results indicate some overfitting, moreover it performs better in the Hit-rate metric than the city-district hedonic model. Note that the sampling is conducted once and there might some variance in the sample estimates, due which observations constituting which group.

A.4 Description of variables

Table 17: Description of variables

Name of variable	definition / Norwegian translation	mean(std.dev)
totalprice(P)	salesprice+commondebt	3 690 068(1 427 243)
salesprice	the final selling price excluding commondebt	3 548 322(1 485 244)
commondebt(CD)	Fellesgjeld	140 632(298 034)
Size(SI)	P-rom(Livingarea)	66.35(22.79)
numbedroom(NB)	Number of bedrooms	1.79(0.75)
floor(F)	floor of the apartment	2.99(1.86)
buildyear(BY)	Year the building was built	1959(36.97)
year	year of the transation	2015(0.82)
month	month of the transaction	6(3,32)
garage(G)	dummy which =1 if parking lot is included in the transaction	0.33
coop(CO)	dummy which=1 if apartment is a part of a housing cooperative(borettslag)	0.45
balcony(B)	dummy which=1 if the header of the advert includes the Norwegian word "balkong"	0.50
aview(V)	dummy which=1 if the header of the advert includes the Norwegian word "utsikt"	0.16
fireplace(FI)	dummy which=1 if the header of the advert includes the Norwegian word "peis"	0.09
needsrenovating(NR)	dummy which=1 if the header of the advert includes the Norwegian word "oppussing"	0.05
renovated(R)	dummy which=1 if the header of the advert includes the Norwegian word "nyoppusset"	0.03
sunny(SU)	dummy which=1 if the header of the advert includes the Norwegian word "solrik"	0.22
heatingrating(HR)	the colourvalue value"green to red" of the energimerke score of an aparment where 1 equals green and 5 equals red, it describes how much of the heating stems from other sources then electricity and fossil fuels, with 1 being large parts and 5 being very little	3.76(1.52)
energyrating(ER)	the lettervalue of the energimerke score of an apartment where 1 equals A and 7 equals G. It describes how energy effecting the apartment is, with A being best and G being worst	5.65(1.42)
citydistrict(CD)	the city-district of the apartment	