Discussion Papers
Statistics Norway
Research department

No. 865
October 2017

Erlend Eide Bø

Match quality in housing transactions:
What can we learn from comparing buyers and sellers?

Statistics Norway
Statistisk sentralbyrå

Discussion Papers No. 865, October 2017
Statistics Norway, Research Department

Erlend Eide Bø

Match quality in housing transactions:
What can we learn from comparing buyers and sellers?

Abstract:
Match quality, the part of housing value to the buyer which is unique for each buyer-house match, is important in several housing market matching models, but measuring it is difficult for an econometrician. I suggest that similarity between buyers and sellers (at the time they bought) may be used to measure match quality. Successive owners of houses should share characteristics if observable characteristics of a buyer are correlated with the buyer's preferences for housing. A buyer could expect to have a high match quality if similar to the seller. I use a simple matching model to show this mechanism. I test this prediction using unique data with information on buyers and sellers (at the time they bought), and show that their similarity can be used as a proxy for match quality. Buyers who resemble sellers are paying more, also when a large number of observable housing characteristics are controlled for.

Supplementary analyses strengthen my claim that the distance between seller and buyer can be used as a proxy for match quality. Matches with low distance lead to slightly reduced hazard rate of reselling the house, and an increased probability of having children, both of which would be expected in a high quality match.

Keywords: Taxation, Distribution, Housing

JEL classification: D83, R31

Acknowledgements: I thank Marcus Hagedorn and Edwin Leuven for help and advice. Kjetil Telle and participants at Forskermøtet 2016 provided valuable comments.

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

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

© Statistics Norway
Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no/en/forskning/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

I denne artikkelen forsøker jeg å finne et mål på hvor godt bolig og boligeier passer sammen. Dette er vanskelig observerbart, men å ha et godt mål kan være nyttig av flere årsaker. For det første gir mange boligmodeller prediksjoner på hvordan dette målet samvarierer med andre, observerbare kjennetegn ved boligmarkedet, og målet kan dermed brukes til å teste hvorvidt disse prediksjonene holder. I tillegg kan et mål på hvorvidt eier og bolig passer sammen også brukes til å studere for eksempel hvor viktig villighet til å flytte er for varigheten av arbeidsløshet.

Ideen bak målet er at hvorvidt en kjøper ligner på forrige kjøper av samme bolig inneholder informasjon. Enhver person vil verdsette en bolig forskjellig, fordi vekten som legges på geografisk plassering, planløsning, standard o.l. varierer. Hvis man antar at denne verdsettingen av bolig er korrelert med observerbare kjennetegn ved personen (for eksempel alder og utdanning), så vil nye boligkjøpere passe godt med boligene de kjøper hvis de er like de forrige kjøperne. Dette vises teoretisk i en enkel søkemodell.

Deretter bruker jeg detaljert informasjon om boligkjøper og boliger over perioden 2007 – 2014 til å konstruere et mål på likhet mellom nye og forrige kjøpere. Likhet måles ved alder, kjønn, husholdningsstørrelse, antall barn, og utdanningsnivå.

Kjøperpar er likere i folkerike kommuner enn de er i kommuner med få innbyggere. Og i perioder med høye boligpriser er kjøperpar likere enn i perioder med lave priser. Dette stemmer med teorien om at det er lettere å finne en passende bolig i markeder med mange salgsobjekter.

Ved hjelp av hedonisk regresjon viser jeg at kjøpere som er like de forrige kjøperne betaler en høyere pris enn de som ikke er like. Dette gjelder også hvis man kontrollerer for et stort antall kjennetegn ved boligene. At like kjøpere betaler mer betyr at de verdsetter boligen mer, og støtter ideen om å bruke likhet mellom kjøper og forrige kjøper som mål på hvor godt bolig og boligeier passer sammen.

I tillegg styrkes bruken av likhet mellom kjøper og forrige kjøper som mål ved at like kjøpere blir boende lenger i boligene sine, og at de i større grad får barn i årene som følger kjøp.

1 Introduction

Match quality in a housing market setting is the unique buyer valuation which exists for each buyer-house match. It is separate from the common utility of the house, which can be seen as the average value of the amenities of the house. The idea is that even though all buyers see the same house, each buyer will put an unique value on the house's combination of location, spatial layout, view and so forth. In housing search models, match quality is mostly modeled as a random variable, which is unobservable to observers.

Being able to measure match quality would be useful to evaluate the housing search models that use the concept, as they often include predictions for how match quality differs over states of the economy. It would also allow for further research into what effects housing match quality has on outcomes such as e.g. mobility. There has been suggestions that the reduced mobility of home-owners leads to higher unemployment.[1] If that is the case, not only ownership percentages, but also match quality levels may affect unemployment.

The concept of match quality is important in several housing market models.[2] Anenberg and Bayer (2013) develop a housing search model with an important role for the timing of buying and selling of those agents who want to both buy and sell a house. Agents' choice of selling before buying or buying before selling is shown to amplify the volatility of housing market cycles. Match quality is normally distributed and idiosyncratic to each meeting between buyer and house. Even though the distribution of match quality in meetings is similar between periods, the dynamics of the model mean that the average match quality for housing transactions differ between periods. In a seller's markets, when there are many buyers compared to sellers, agents require a lower draw of match quality to buy before selling than in a buyer's market.

The focus in Ngai and Tenreyro (2014) is on how the thick-market effect leads to seasonal housing markets. Housing markets show strong, predictable seasonal patterns, where "hot" markets with high prices and numerous transactions alternate with colder markets. To explain this, Ngai and Tenreyro model a housing market with thick-market effects, i.e. where the expected match quality is positively correlated with the number of houses for sale. They also show empirical support for the idea that average match quality is higher for houses transacted in hot markets than in cold markets.[3] The thick-market effect amplifies small differences in moving probability into sizable seasonal trends.

Expanding on the thick-market effect, Nenov et al. (2016) show that greater heterogeneity in the housing stock amplify the thick-market effects. This would lead to a prediction that the effect of match quality on prices would be lower in areas with relatively homogeneous

[1] See e.g. Head and Lloyd-Ellis (2012), who find a link between home-ownership and unemployment. The effect is small when the model is calibrated to match real data.

[2] Match quality also plays a central role in the literature of labor market matching (e.g. Pissarides, 1985; Bowlus, 1995) and marriage matching (Weiss and Willis, 1997).

[3] They proxy match quality by the duration of stay, and by the number and cost of repairs and alterations performed on the house in the first two years of ownership.

housing. The match quality of current housing owners could also affect the number of housing transactions in the future, as in the model of Ngai and Sheedy (2016). There, the moving decision depends on the match quality. All housing owners with a match quality below a cut-off dependent on the state of the economy choose to move.

However, while match quality is usually assumed observable for the housing buyer, it is not easily observable for an econometrician. In housing search models, match quality is mostly modeled as a random variable. Intuition would perhaps suggest measuring match quality as the residual of price from a hedonic price regression, but that measure may be distorted by the heat of the housing market. In hot housing markets, with the increased possibility of bidding wars (Han and Strange, 2014), a high price may represent high bargaining power for the seller rather than a high match quality.[4] The same concern would apply to using time-on-market as a proxy for match quality. The perhaps most obvious measure of match quality is how long a buyer actually remains in a house. But this measure requires collecting data for a number of years following purchase. The measure suggested here allows for an almost contemporaneous measure.

This paper presents a novel idea for how to use rich data to measure match quality. Assuming observable characteristics of buyers are correlated with their preferences for housing, successive owners of houses should share characteristics. An observer could expect a new owner to have higher match quality if similar to the old owner than if dissimilar. I develop a simple matching model based on Ngai and Tenreyro (2014) to show this mechanism.

In the empirical part of the paper, I connect information on present sellers from the time when they bought ("previous buyers"), with characteristics of the present buyers. The reason for using the information on sellers from the time they bought instead of the time they sell is that the I want to measure how similar the two buyers are. Imagine a 30 year old buyer, who finds a house she likes, stays in a house for 10 years, and gets 2 children, before she needs more space and decides to sell. I would expect the next buyer to be well matched if she is 30 years old, without children, not if she is 40 with two children.

Using the connected buyer characteristics it is possible to measure the similarity between the buyer and the previous buyer (who is now selling). I show that the (Mahalanobis) distance between buyer and previous buyer can be used as a proxy for match quality. Buyers who resemble the previous buyers are paying more, also when a large number of observable characteristics are controlled for. Note that I do not claim that this measure can be used to evaluate match quality for every single transaction, but the measure is useful for finding the average match quality, e.g. in a certain period, or equivalently, the probability that a match is good.

Supplementary analyses strengthen my claim that the distance between seller and buyer can be used as a proxy for match quality. Matches with low distance seem to survive

[4] See i.e. Carrillo (2013) for a model where high prices are a result of the bargaining power of the seller.

longer. Low distance matches also lead to an increased probability of having children, which I argue would be expected in a high quality match. The distance metric is clearly significant in a logit regression of increased number of children in the household in the years after the house is bought.

In Section 2, a matching model is presented. The data I use is described in Section 3, followed by an explanation of the distance metric in Section 4. Section 5 contains the results, while Section 6 concludes.

2 Model

In this section, a simple search and matching model is sketched. The purpose of the model is to illustrate how the observable characteristics of buyers and sellers can be used to measure match quality. This is done to explain why I interpret the similarity of buyer and past buyer shown in Section 5 as a proxy for match quality.

The model is based on a non-seasonal version of the model in Ngai and Tenreyro (2014).[5] The main new feature of the model is that the match quality depends on heterogeneous types of buyers and houses. Match quality does not only depend on a randomly drawn value, as in Anenberg and Bayer (2013) or Ngai and Tenreyro (2014), but also reflects that some observable characteristics of a buyer (e.g. age, number of children) increase match quality for certain kinds of houses.

The economy consists of a unit measure of infinitely lived, risk neutral agents who receive utility from owning a house. The agents have three states: owners, buyers and sellers.

A measure o of agents are owners, who are matched with their houses. They receive utility per period for being matched, which depends on the individual match quality.

Each period, the probability of a mismatch shock is δ, in which case the agents become sellers. The house is put for sale, and per period utility is u. After sellers have sold their house, they receive utility equal to the transaction price of the house, and exit the economy.

Buyers enter the economy at rate δo, keeping the population constant. Buyers meet sellers in a market with search frictions.

So far, these are standard assumptions. In addition, owners and buyers differ over a set of characteristics X, the distribution of which is similar at all times.[6] To clearly expose the mechanics of the model, I will model X as a single variable with two possible values $[x_1, x_2]$, thus there are two types of buyers in the model, with measure b_1 and b_2.[7] Similarly, there are two types of houses for sale, with measure v_1 and v_2, one that is (slightly) preferred by buyers of type 1, the other by type 2.

[5] Though unlike Ngai and Tenreyro, there is no thick-market effect on match quality in my model.

[6] I assume that it is impossible for buyers to observe the sellers X.

[7] This is done for simplicity in presentation. Generalizing X to a more flexible distribution would give similar results.

The per period returns from a house can be modeled as $u + \epsilon_i$, where u is the flow utility from owning the house common to all prospective owners. It can be seen as the average value of the house's amenities. Match quality, ϵ_i, is individual and unique for each match between house and house owner. I assume that match quality is uncorrelated with u, and correlated with the observable characteristics of the buyer, or buyer type: $\epsilon_i = \gamma_{jk} + \eta_i$, where $j \in [b_1, b_2]$ and $k \in [v_1, v_2]$. Thus, the match value for a house is a common valuation, u, plus the match quality, which is a function of how well X (i.e. buyer age) fits with the house type, and a stochastic error term η.

The value of a buyer of type j buying a house of type k is thus:

$$H(b_j, v_k, \eta) = u + \gamma_{jk} + \eta + \beta[(1-\delta)H'(b_j, v_k, \eta) + \delta V'(v_k)], \tag{1}$$

where γ_{jk} is higher if $j = k$, and η is i.i.d. and drawn from the distribution $F(\eta)$. The value function of a particular match $H(b_j, v_k, \eta)$ is the present period value of the match, plus the value of the match in next period, adjusted by the probability of a moving shock, δ, occurring in next period. The discount rate is given byβ.

The match quality of a house can only be observed during a visit, so ex-ante, buyers know neither the type of a house nor the stochastic match quality (but they do know the distribution of houses for sale).[8] If house type was known in advance, all buyers of one type would buy houses of "their" type.

The total surplus when a seller and buyer meet is:

$$S(b_j, v_k, \eta) = H(b_j, v_k, \eta) - \beta(B'(b_j) + V'(v_k)) + u, \tag{2}$$

where $\beta(B'(b_j)+V'(v_k))$ is the discounted value of remaining respectively a buyer of type j and a seller with a house of type k in next period

I assume that the draws of γ and η are common knowledge during a meeting between buyer and seller. A transaction thus happens if the surplus of a meeting is positive, $S(b_j, v_k, \eta) \geq 0$, or using (2) :

$$H(b_j, v_k, \eta) - \beta(B'(b_j) + V'(v_k)) + u \geq 0. \tag{3}$$

What can be observed in (3) is that the surplus of a meeting depends on the value of the match, H. As expected H is higher if buyer and house type correspond, so is the meeting surplus. Prices are determined as a bargaining problem between buyer and seller. I model the bargaining process as Nash bargaining with weights θ and $(1-\theta)$ for seller and buyer respectively, but the specific bargaining process is not important. Any bargaining process with a bargaining weight higher than 0 for the buyer will give higher prices when

[8]The buyer observes the match quality fully during a visit. Unlike the labor market search model of Jovanovic (1979), there is no learning about match quality over time.

the surplus is higher and be sufficient for the results below: prices are higher when a buyer is of the same type as the past buyer.

2.1 Steady state equilibrium

I have shown that the expected matching surplus and price are higher when the buyer and house types match. What I want to show is how prices vary with the match between buyer type and past buyer type. To look at that dynamic, I here solve the model for steady state. In steady state, the owner value function becomes:[9]

$$H_{jk}(\eta) = \frac{u + \gamma_{jk} + \eta}{1 - \beta(1-\delta)} + \frac{\delta V_k}{1 - \beta(1-\delta)}. \tag{4}$$

Transactions occur if $\eta_i \geq \eta^*_{jk}$, where η^*_{jk} is the lowest draw of the idiosyncratic match quality which makes the surplus non-negative:

$$\eta^*_{jk} =: H_{jk}(\eta) = \beta(B_j + V_k) + u. \tag{5}$$

Using the definition from (5) in the owner value function (4) gives:

$$\gamma_{jk} + \eta^*_{jk} = (1 - \beta(1-\delta))\beta B_j + ((1 - \beta(1-\delta))\beta - \delta)V_k - \beta^2(1-\delta)u. \tag{6}$$

From (6) it can be noted that with a match between buyer and house type (γ_{jk} high), the idiosyncratic match draw needed for a transaction, η^*_{jk}, is lower than if there is a mismatch between buyer and house type.

By using (5) it is also possible to rewrite the surplus of a match:

$$S_{jk} = H(b_j, v_k, \eta) - H(b_j, v_k, \eta^*_{jk}) = \frac{\eta - \eta^*_{jk}}{1 - \beta(1-\delta)}, \tag{7}$$

which means that the expected surplus for a match that leads to a transaction can be written as $E(S_{jk}|\eta > \eta^*_{jk}) = \frac{E[\eta - \eta^*_{jk}|\eta > \eta^*_{jk}]}{1-\beta(1-\delta)}$

The value function of a buyer of type j is

$$B_j = \beta[B'_j + (1-\theta)\sum_k (\frac{v_k}{v^t}(1 - F(\eta^*_{jk}))E[S_{jk}|\eta > \eta^*_{jk}])] \tag{8}$$

where the total measure of houses, $v^t = \sum_k v_k$, and $\sum_k \frac{v_k}{v^t}(1 - F(\eta^*_{jk}))$ is the probability for a buyer of type j that a transaction goes through. The buyer gets the continuation value of being a buyer in next period (B'_j), plus a share $(1-\theta)$ of the surplus of a transaction

[9] To save space, I use subscripts j, k in the following to denote that value functions depend on buyer and/or house type.

if it goes through. Both the surplus and probability of a transaction depends on the type of house that the buyer inspects. As the buyer cannot in advance observe the type of house, the probability of the buyer of meeting a seller with house k depends on the share of houses for sale of each type.

Using (7), the buyer value function can be written as:

$$B_j = \beta[B_j' + (1-\theta)\sum_k(\frac{v_k}{v^t}\frac{h^*(\eta_{jk}^*)}{1-\beta(1-\delta)})], \tag{9}$$

where $h^*(\eta_{jk}^*) = (1-F(\eta_{jk}^*))E[\eta - \eta_{jk}^*|\eta > \eta_{jk}^*]$ is the expected surplus of a match.

In steady state:

$$B_j = \frac{1-\theta}{(1-\beta)(1-\beta(1-\delta))}\sum_k(\frac{v_k}{v^t}h^*(\eta_{jk}^*)) \tag{10}$$

The value function of a seller with house of type k is

$$V_k = u + \beta[V_k' + \theta\sum_j(\frac{b_j}{b^t}\frac{h^*(\eta_{jk}^*)}{1-\beta(1-\delta)})], \tag{11}$$

where $b^t = \sum_j b_j$. The seller value function is quite similar to the buyer value function. The seller gets the continuation value of being a seller in next period, V_k', plus a share of the surplus of a transaction if it goes through, which depends on the type of buyer that visits the house. In addition the seller gets the value u of owning a mismatched house.

The steady state seller value function is:

$$V_k = \frac{u}{1-\beta} + \frac{\theta}{(1-\beta)(1-\beta(1-\delta))}\sum_j(\frac{b_j}{b^t}h^*(\eta_{jk}^*)) \tag{12}$$

The law of motion for mismatched houses (or houses for sale) of type k is

$$\begin{aligned} v_k' &= \delta(\sum_j[\frac{b_j}{b^t}v_k(1-F(\eta_{jk}^*))] + 1 - v_k) + \sum_j[\frac{b_j}{b^t}v_kF(\eta_{jk}^*)] \\ &= (1-\delta)\sum_j[\frac{b_j}{b^t}v_kF(\eta_{jk}^*)] + \delta. \end{aligned} \tag{13}$$

The first term in the first equation are houses which are put for sale in the current period, a share δ of the houses which were not previously for sale. The second term represents the unsold houses from the last period, which are the houses where the η drawn in the match were too low for a transaction to occur. In steady state, the law of motion for houses can be written as:

$$v_k = \frac{\delta}{1-(1-\delta)\sum_j\frac{b_j}{b^t}F(\eta_{jk}^*)}. \tag{14}$$

Similarly, the steady state law of motion for buyers of type j is

$$b_j = \frac{\delta}{1-(1-\delta)\sum_j \frac{v_j}{v^t} F(\eta^*_{jk})}. \tag{15}$$

The expected price of a transaction where a type j buyer transacts with a seller of type k is the solution to a Nash bargaining problem over the surplus of the transaction:

$$E[P_{jk}] = (1-\theta)\frac{u}{(1-\beta)} + \theta E[H(X_j, v_k, \eta)|\eta > \eta^*_{jk}]. \tag{16}$$

Using the fact that $E[H(X_j, v_k, \eta)|\eta > \eta^*_{jk}]$ can be rewritten as $H(X_j, v_k, \eta^*_{jk}) + E[S_{jk}|\eta > \eta^*_{jk}]$, the price is given as (see Appendix A for details):

$$\begin{aligned} E[P_{jk}] &= \frac{u}{(1-\beta)} + \frac{\theta}{(1-\beta(1-\delta))}[\beta\frac{1-\theta}{(1-\beta)}(\sum_k[\frac{v_k}{v^t}h^*(\eta^*_{jk})]) + \\ &\quad \frac{\theta}{(1-\beta)}(\sum_j[\frac{b_j}{b^t}h^*(\eta^*_{jk})]) + E[\eta - \eta^*_{jk}|\eta > \eta^*_{jk}]] \end{aligned} \tag{17}$$

2.2 Theoretical results

For a house of a certain type k, the share of buyers j with X_j giving a high γ_{jk} should be higher than their share in the population, even though the matching of houses and buyers is random. This result is due to the fact that $Pr[S(\epsilon_i) \geq 0]$ increases with γ_{jk}, as there is less need for a high draw of the random match quality η.

In steady state, this result will hold for each period. Thus, a disproportionate share of sellers will also be of a type that matches the house. This means that the probability of a transaction is high when the seller used to be the same type as the buyer.

Secondly, the match quality of buyers involved in matched transactions ($j = k$), should on average be higher than that of other buyers. Higher average match quality will be reflected in higher prices. The expected surplus $E[S_{jk}|\eta > \eta^*_{jk}]$ of a transaction also increases with γ_{jk} as the average match value increases. This can be seen by combining (6) and (7).

$$S_{jk} = \frac{\eta + \gamma_{jk} - (1-\beta(1-\delta))\beta(B_j + V_k) + \delta V_k - \beta^2(1-\delta)u}{1-\beta(1-\delta)} \tag{18}$$

When the surplus is high, bargaining over prices means that the price is also high.

The size of these effects can be shown in numerical simulations.

2.3 Numerical results

I simulate the model using parameter values taken from Ngai and Tenreyro (2014). Each period represents half a year. The implied yearly interest rate β is equal to 6 percent and the yearly user cost of housing u is 3 percent of the housing price.[10] The rate of moving shocks, δ is set to get an average duration of stay of 13 years, while the bargaining weight of sellers, θ, equals 0.5. The model is symmetric, i.e. the share of buyers and houses of both types is 0.5, and the value of γ (the preference of a buyer for a house of same type) is similar for both buyer-house type match.

Table 1 presents model simulations of the share of buyers that buy houses of their favored type in column (2). It then shows the price they pay relative to the price of houses bought by buyers of the other type, as the value of γ for matched buyer house types relative to mismatched buyer house types changes. In all simulations, the value of γ for mismatched types ($j \neq k$) is 0. The first row is the case where there is no difference in preferences between types ($\gamma = 0$). The table indicates that as γ for $j = k$ increases, the share of buyers buying from sellers who used to be the same type is increasing. Also, the higher γ, the larger the price mark-up those buyers pay compared to buyers of the other type.

Table 1: Numerical results

$\gamma(j=k)$	Share buyers j=k	Price premium j=k	Share buyers j=j-1 where j=k	Share buyers j≠j-1 where j=k	Price premium j=j-1
0.00	0.500	1.000	0.500	0.500	1.000
0.05	0.514	1.003	0.529	0.500	1.000
0.10	0.529	1.006	0.557	0.500	1.000
0.15	0.543	1.008	0.585	0.500	1.001
0.20	0.557	1.011	0.613	0.500	1.001
0.25	0.571	1.014	0.640	0.500	1.002
0.30	0.585	1.017	0.666	0.500	1.003
0.35	0.600	1.020	0.692	0.500	1.004
0.40	0.614	1.022	0.717	0.500	1.005
0.45	0.628	1.025	0.741	0.500	1.006
0.50	0.643	1.028	0.764	0.500	1.007

Notes: Results of numerical simulations of the model for different levels of type preference (γ).

For comparisons with the empirical part of this paper, it is interesting to calculate the properties of transactions between types of buyers and previous buyers. Column (4) and (6) of Table 1 reports the share of buyers buying from similar past buyers, and the price they pay compared to those who do not buy from similar past buyers.

To explain the calculation of these results, I use the numbers from the last row as an example. In a steady state, the past distribution of sellers and buyers is similar to the

[10] A model where all buyers are of the same type, and all houses supply utility $u + \eta$ to any matched owner is used to calibrate the values of u (and the η^* and P needed to find u).

present distribution. Looking at those buying a house of type 1, the share of type 1 buyers buying from past type 1 buyers is $0.643 \times 0.643 = 0.413$. A smaller share, 0.230 are type 1 buyers buying from type 2, similarly, 0.230 are type 2 buyers buying from type 1. Lastly, 0.127 are type 2 buyers buying from type 2. Thus, 76 percent of the buyers buying from similar buyers are matched with their type of house, versus 50 percent of those buying from different buyers. The average price buyers that are similar to past buyers pay is higher than for non-similar buyers (1.007 times the non-match price). This increase in price, for buyers who buy from similar past buyers, is what I look for in the empirical part of the paper. The size of the price premium is dependent on a lot of assumptions in the simulations, and is thus not so interesting in itself.

3 Data

Housing transaction and ownership data from three different sources are merged with a register covering personal characteristics. Data on registered transfers of real property (*Tinglysning)* come from the Norwegian Mapping Authority, and include the personal identifier of sellers and buyers, as well as some information on the transacted house. The data cover the period 2007-2014, though transactions of co-operative housing are not included. Additional data on transactions are gathered from the main Norwegian housing search web page (Finn.no) which covers a large share of market transactions. These data include information on transaction price, time on market and housing characteristics, including appraisal value.

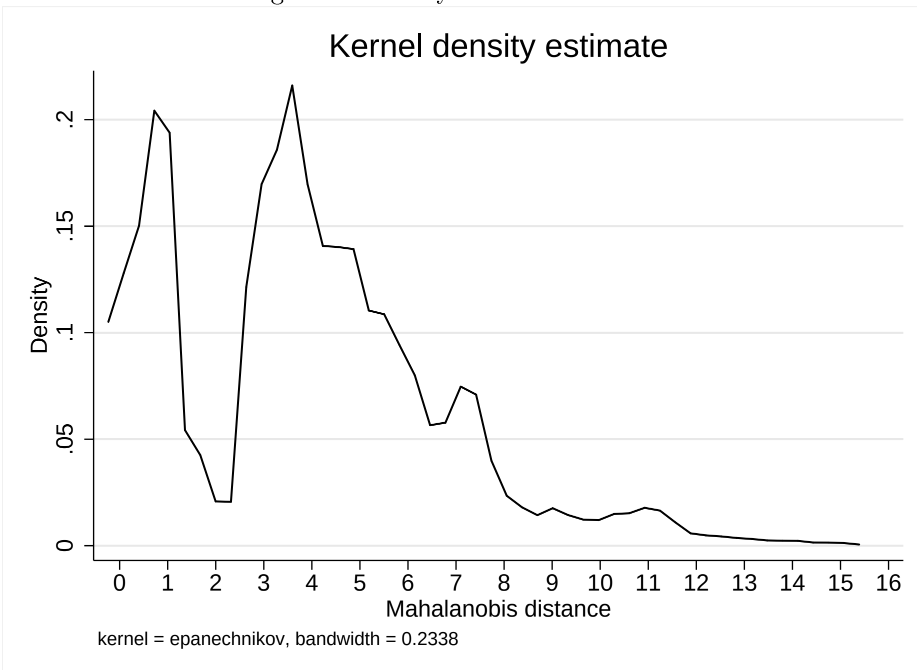
A third data source is the Norwegian cadastre (*Matrikkelen)*, which holds information about ownership of all housing, at least back to 2004. Through the cadastre, length of ownership is found, also for ownerships where I do not have repeated transactions in the time period 2007-2014. Since a starting year of 2004 in the data may reflect either a transaction taking place in 2004 or the ownership being registered in the cadastre, I am only able to use repeated transactions taking place in the period 2005 to 2014.

The data on registered transfers, housing sales and ownership history are connected through housing registry information which not only allows for identification of single houses, but also units of multi-unit housing. Through the personal identifier, personal characteristics of both seller (at the time the seller bought the house) and buyer is added: income, wealth, level of education, previous house ownership, household size and number of children.[11] The information on personal characteristics comes from the Income and wealth statistics for households (Statistics Norway, 2015), which covers the whole population of Norway and includes information from income tax returns, education registers, etc.

All in all, the data set covers 139,688 repeated sales; pairs of housing transactions where I know the characteristics of the buyers in both set of transactions, and have informa-

[11] Information is from December 31st of the year prior to buying

Figure 1: Density of the Mahalanobis distance



Notes: This figure plots the density of the estimated Mahalanobis distance.

tion on transaction price in at least the second transaction. Data on some of the other characteristics, in particular the appraisal value of houses, are limited to a smaller sample.

4 The distance metric

In the literature on assortative mating, which this paper somewhat resembles, equality between spouses is measured in terms of e.g. their education levels (Mare, 1991) or occupation (Kalmijn, 1994). As for the housing market, other household characteristics, such as age and household characteristics (size and number of children) also seem important determinants of housing demand. When including age and household characteristics, the one-dimensional measures of similarity normally used for quantifying assortative mating can no longer be applied.

In this paper, the similarity between buyer and previous buyer is measured as the Mahalanobis distance (Mahalanobis, 1936). The Mahalanobis distance measures the multivariate distance between observations, scaled by the covariance of covariates (Mardia et al., 1989). If the covariance matrix is the identity matrix, it is equal to the Euclidean distance. The variables used to measure similarity are age, household size, number of children under 18, and dummies for no high school, high school and university education. I drop the observations above the 99th percentile of the Mahalanobis distance to avoid outliers influencing the results.

The distribution of the distance metric is bi-modal, as shown in Figure 1, a first peak where the education of buyer and previous buyer is similar, the second where the education is different. The mean and median value of the distance is 3.8 and 3.6 respectively.

Table 2: Mean Values

	(1) All	(2) Similar	(3) Dissimilar	(4) T-test
Wage	391,998	408,453	375,543	-15.18
Disposable income	349,067	359,617	338,517	-10.30
Capital income	31,644	34,283	29,005	-2.69
Past housing value	264,435	280,751	248,118	-13.40
Financial wealth	744,508	768,297	720,718	-0.78
Age buyer	38.4	38.6	38.2	-5.00
Household size	2.51	2.30	2.71	59.33
Children under 18	0.61	0.50	0.72	45.48
Share below high School	0.19	0.08	0.31	108.19
Share high School	0.38	0.42	0.35	-27.25
Share university	0.42	0.50	.35	-57.22
Buying year	2010.6	2010.7	2010.6	-3.58
Transaction price	2,636,053	2,784,141	2,487,968	-38.53
Housing size	107.5	106.2	108.7	8.40
Share villa	0.38	0.36	0.41	21.43
Share flat	0.46	0.49	0.43	-19.82
Observations	137,239	68,619	68,620	

Notes: Descriptive statistics for the data set used for analyses. Column (1) presents all observations, column (2) and (3) the observations with Mahalanobis distance respectively below and above the median. Column (4) shows a T-test of equality between (2) and (3).

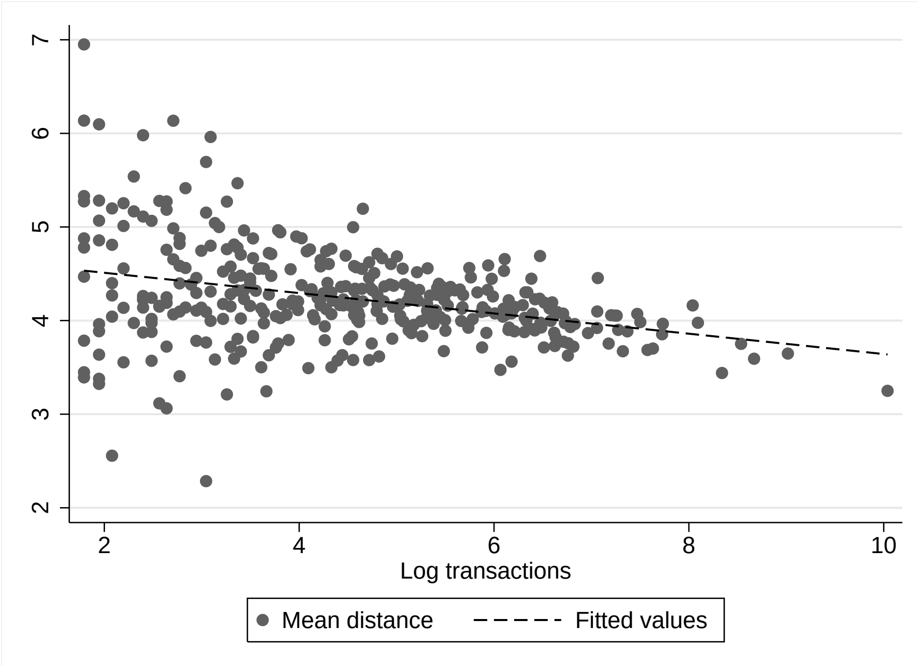
Table 2 gives the descriptive statistics of the observations with a distance below and above median. In similar matches buyers are more educated and have somewhat higher income. Similar matches also occur at higher prices even though the housing size is slightly smaller, which suggests either higher quality houses or more attractive locations. There are also fewer villas, but more flats in the similar matches group. The table also shows a T-test of similarity between the groups. Most variables are significantly different.

It is possible that cities, with larger housing markets and more heterogeneous housing stock, have more segmented housing markets, and thus more similar buyers.[12] In the results, I will use specifications with municipality fixed effects to control for this possibility.

The model of Ngai and Tenreyro (2014) predicts higher match quality in seasons with thick markets. Similarly, one would expect better matches in bigger cities, where the market is thicker. This is confirmed in Figure 2, which plots the mean of the Mahalanobis distance against the log number of transactions in each municipality.[13] There is a clear trend for a large number of transactions (a thick market) being correlated with low distance (more similar matches). The low distance for larger municipalities also holds consistently

[12]The segmentation of housing markets in cities is explored in Piazzesi et al. (2015).
[13]A few municipalities with only one transaction not displayed.

Figure 2: Mahalanobis distance, large and small markets



Notes: This figure plots the average Mahalanobis distance against the log number of transactions within each municipality with more than 5 transactions. It also fits a linear regression of the correlation between distance and log transaction numbers.

over time. In Figure 3, transactions from the five municipalities with most transactions,[14] marked "Large cities", is contrasted against transactions from the remaining municipalities ("Other municipalities"). While the monthly trends look quite similar, the level of the distance is clearly lower in the large cities.

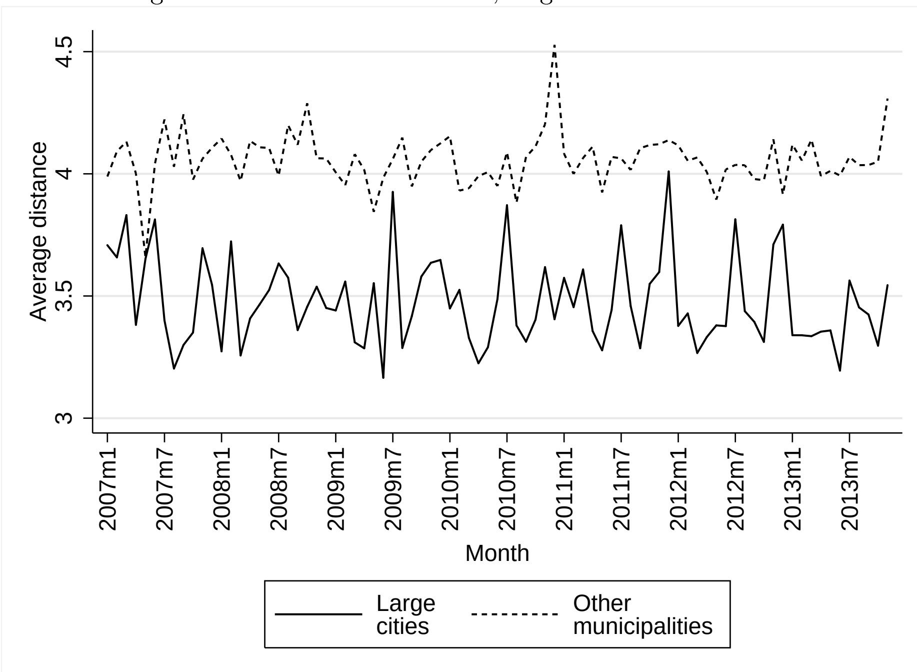
Next, in Figure 4, I show the distribution of the Mahalanobis distance over time, together with the development of housing prices. While the average value of the Mahalanobis distance shows some variation from month to month, it seems like the value is usually low in times with high housing prices. This means that new buyers are closer to old buyers in periods with high prices. A similar pattern holds for the relationship between distance and number of transactions; transaction number is negatively correlated with distance (see Figure A.1 in the appendix).

While there is a clear negative correlation between prices and distance, the changes in this correlation over time also show some interesting patterns. The 12-month rolling correlation between log prices and Mahalanobis distance is shown in Figure 5. The months referred to in the figure are the starting months of each 12 month window.

The correlation is negative over the whole period, but there is a downward trend beginning in mid 2008, continuing until the first part of 2012. Interestingly, this trend appears to

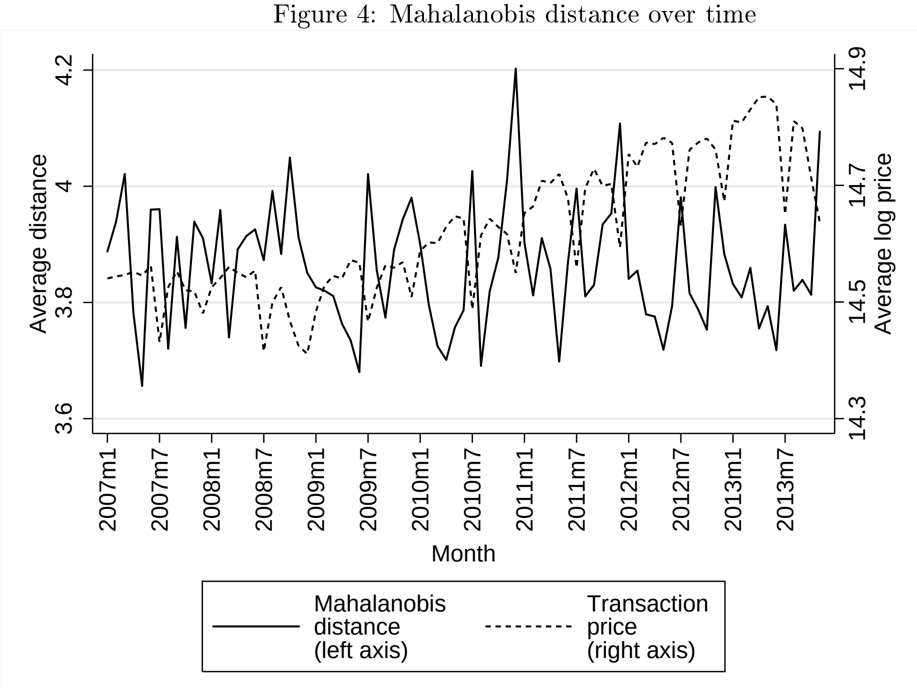
[14]Oslo, Bergen, Trondheim, Stavanger and Bærum.

Figure 3: Mahalanobis distance, large and small markets over time



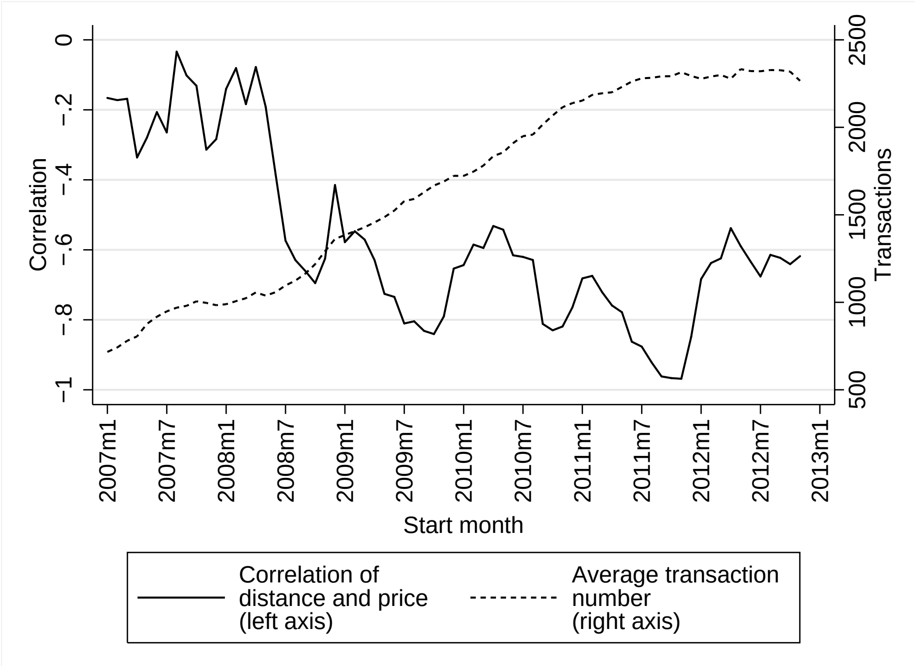
Notes: This figure plots the average Mahalanobis distance of observations in large cities and other municipalities. Large cities are the municipalities of Oslo, Bergen, Trondheim, Stavanger and Bærum.

Figure 4: Mahalanobis distance over time



Notes: This figure plots the average Mahalanobis distance and log housing price over time.

Figure 5: Rolling correlation of Mahalanobis distance and price, and transaction numbers



Notes: This figure plots the 12-month rolling correlation of Mahalanobis distance and price, and the average number of housing transactions over the same 12 months.

coincide with a time when the housing market went from a period of stagnation (following the 2008 financial crisis) to a period of growth, as shown in Figure 5 by the growing average number of transactions over the same period. This, as well as the difference in distance between large cities and smaller municipalities, is in line with the model in Ngai and Tenreyro (2014), where thick-market effects lead to high match quality when the stock of houses for sale is larger.

5 Results

I use the housing transaction data, and information about the Mahalanobis distance between the buyer in the current and the previous transaction to see if the distance correlates with housing price. The regressions presented here are informed by the simple, theoretical search model in Section 3. The dependent variable is the log housing price, while the independent variables are the Mahalanobis distance and log appraisal value (which is a proxy for the quality of the house to the average buyer, u in the model).

The main difference from the model is that I now have to worry about the impact of the housing price cycle, which is not included in the model, and the possibility that match quality differs over housing type and geography. Thus, later specifications add additional controls: month and year fixed effects, variables reflecting the income and wealth of buyers and characteristics of the house, municipality fixed effects, and finally characteristics of the seller. This last specification is shown in equation (19), with log price dependent on the Mahalanobis distance, the appraisal value, a set of buyer, house and seller characteristics, plus time and municipality fixed effects.

$$\ln(p_{imt}) = \beta_0 + \beta_1 md_{imt} + \beta_2 \ln(ap_{imt}) + \beta_3 B_{imt} + \beta_4 H_{imt} + \beta_3 S_{imt} + \mu_t + \gamma_m + \varepsilon_{imt} \quad (19)$$

The main results are presented in Table 3. Through all specifications, the distance coefficient is significant and negative. Buying from a seller who had different characteristics is correlated with lower price, and I interpret that as a sign of lower match quality.

Also worth noting is that the appraisal value is, unsurprisingly, always very important for the price. But when other information is added, in particular municipality fixed effects, the coefficient decreases from above 0.95 to around 0.8. The signs on the remaining coefficients are mostly as expected, with wealth, income, family size and education positively correlated with the price.

The last specification adds seller characteristics. Seller characteristics do matter for the price; both a likelihood ratio test and a Wald test reject the hypothesis that seller characteristics are jointly insignificant. This contrasts with the assumptions from the theoretical model.

It could be imagined that there are differences between housing types which are not fully captured by a dummy variable in the full regression.[15] As a robustness check, Table A.1-A.3 in the appendix show the specifications from Table 3, run separately on row-houses, villas and flats. The size and significance of the distance measures roughly holds for all three specifications, though it is somewhat weaker for villas.

5.1 Ownership length of sellers

Using similarity as a proxy for match quality depends on the idea that past and present buyers should be similar because they are both likely to have an unobserved preference for that particular house. In the data, there is a relatively large amount of housing with a very short ownership length. There is reason to believe that most of these cases are either houses bought, renovated and resold as investments, or houses where the owner feels mismatched straight away. Neither of these cases fit with the theoretical model, where ownership length is only determined by the occurrence of random mismatch shocks.

Here, I look at how ownership length affects the results. Similarity between buyers should not be a predictor of good match quality, and thus excessive price, if the ownership length of sellers has been very short. The observations where the ownership length (i.e. time

[15] Flats are e.g. usually smaller, with higher price per sq.m. and shorter time of stay than other types of housing.

Table 3: Results

Log price	(1)	(2)	(3)	(4)	(5)
Distance	**-0.0183****	**-0.0021****	**-0.0021****	**-0.0014****	**-0.0009****
	(0.0005)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Log appraisal value		0.959**	0.953**	0.802**	0.797**
		(0.0010)	(0.0010)	(0.0018)	(0.0018)
Buyer log Wage				0.0007**	0.0007**
				(0.0001)	(0.0001)
Buyer log Wealth				0.0051**	0.0050**
				(0.0002)	(0.0002)
Size of househ. buyer				0.0013*	0.0011
				(0.0006)	(0.0006)
No. of children buyer				-0.0027**	-0.0028**
				(0.0009)	(0.0009)
High school buyer				0.0099**	0.0098**
				(0.0013)	(0.0013)
University buyer				0.0254**	0.0247**
				(0.0013)	(0.0013)
Villa				-0.0058**	-0.0036*
				(0.0014)	(0.0014)
Flat				-0.0036*	-0.0029
				(0.0015)	(0.0015)
Log square meters				0.113**	0.115**
				(0.0018)	(0.0018)
Size of househ. seller					0.0015*
					(0.0007)
No. of children seller					-0.0004
					(0.0010)
High school seller					0.0137**
					(0.0013)
University seller					0.0237**
					(0.0013)
Month and year FE			yes	yes	yes
Age dummy buyer				yes	yes
Municipality FE				yes	yes
Age dummy seller					yes
Observations	137,239	78,187	78,187	77,856	77,547
R-squared	0.011	0.929	0.931	0.942	0.942

Notes: This table presents the results of OLS-regressions where the dependent variable is the log housing price. Independent variables are the Mahalanobis distance and different control variables. Standard errors in parentheses.

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

between first and second buyer) is less than 12 months are split from observations with ownership length 12 months or above.[16] The specification from column (4) in Table 3 is then run separately for each of the samples, with the results presented in Table 4.[17]

Table 4: Short and longer ownership length

Log price	(1)	(2)
Distance	**-0.00016**	**-0.00157****
	(0.00055)	(0.00017)
Log appraisal value	0.909**	0.789**
	(0.0052)	(0.0019)
Buyer log Wage	0.00027	0.00066**
	(0.00039)	(0.00014)
Buyer log Wealth	0.0042**	0.0052**
	(0.00075)	(0.00024)
Size of househ. buyer	-0.0016	0.0016*
	(0.0020)	(0.0006)
No. of children buyer	0.0068*	-0.0040**
	(0.0029)	(0.0009)
High school buyer	0.0040	0.0102**
	(0.0040)	(0.0013)
University buyer	0.0134**	0.0261**
	(0.0042)	(0.0014)
Villa	0.0116*	-0.0052**
	(0.0045)	(0.0015)
Flat	-0.0128*	-0.0023
	(0.0051)	(0.0016)
Log square meters	0.027**	0.122**
	(0.0057)	(0.0019)
Month and year FE	yes	yes
Age dummy buyer	yes	yes
Municipality FE	yes	yes
Observations	7,586	70,101
R-squared	0.963	0.941

Notes: OLS-regressions, the specification used in Table 3, column (4), run separately on observations with ownership length below and above 12 months respectively. Standard errors in parentheses.
** $p<0.01$, * $p<0.05$

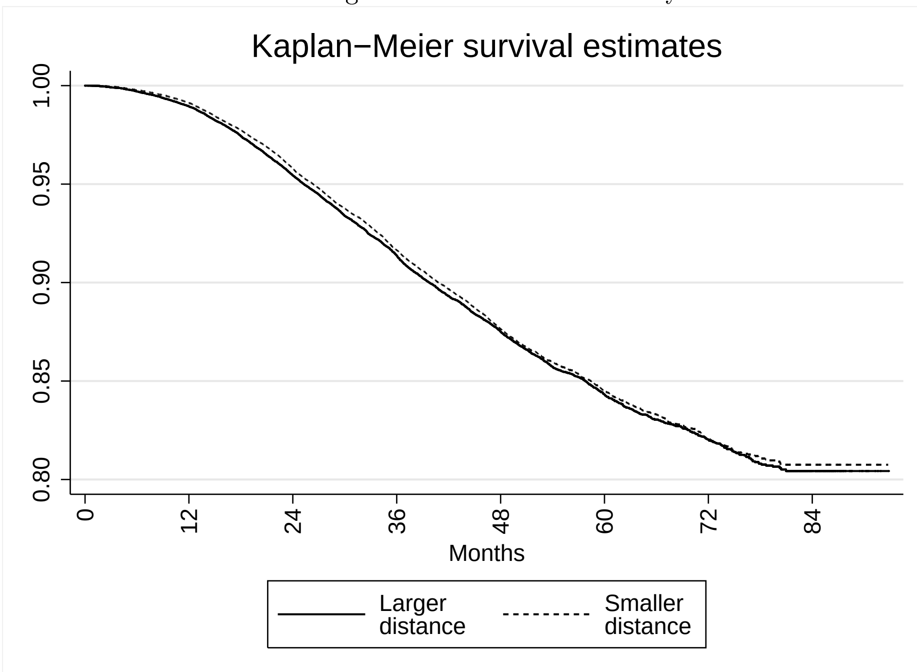
Mahalanobis distance has no significant implications for the price of houses being owned for a short time. This fits with my interpretation of Mahalanobis distance as match quality, as buying from a seller who did not buy due to the quality of the match should offer no predictions on match quality.

Another way to look at the connection between ownership length and my match quality measure is by interacting the two. As shown in Table A.4, in the appendix, the

[16] I do not know the exact date of ownership for many of the first buyers. The start of ownership is set to January 1st of the transaction year.

[17] The average Mahalanobis distances for the samples are respectively 5.82 and 4.65.

Figure 6: Survival rates of buyers



Notes: This figure plots the Kaplan-Meier survival estimates of the buyers with Mahalanobis distance respectively above and below median.

Mahalanobis distance has a larger impact on price with longer ownership length of the seller.

5.2 Consequences of match quality

There are a few likely consequences of having a good housing match. Here, I see if my distance measure are correlated with these outcomes in a way that match quality should be.

The most obvious outcome of having a good match is that expected time of stay should be higher.[18] Unfortunately, I do not observe the buyers for very long following the transaction. Therefore, it is challenging to study whether distance is correlated with the length of stay. But roughly eight percent of the buyers subsequently sell their house during my observation period. I split the sample in two groups, with low and high Mahalanobis distance, and do survival analysis of the two groups. As shown in Figure 6, a slightly higher share of the group with low Mahalanobis distance remains in their homes than the high Mahalanobis distance group, though the difference is not significant.

As shown in the descriptive statistics (Table 2) the group with lower distance are younger

[18]Indeed, in labor models, job tenure is often used as a measure of match quality, see e.g. Bowlus (1995) and Centeno (2004). While the simple model presented earlier in this paper does not display this correlation, the housing search model of e.g. Ngai and Sheedy (2016) does.

and have higher education and income. To see whether the differences in the distribution of housing and buyer characteristics between low and high distance buyers have an impact, I run a Cox proportional hazards model. The results, in Table 5, display some indications that buyers with low distance remain in their home for a longer period. When controlling for observable characteristics of the house, and also when adding municipality fixed effects, the effect of distance is positive on the probability of selling at the 5% significance level.

Table 5: Hazard of selling

	(1)	(2)	(3)
Distance	**-0.0019**	**0.0087***	**0.0076***
	(0.0035)	(0.0038)	(0.0038)
Log price	-0.612**	-0.207**	-0.294**
	(0.0203)	(0.0292)	(0.0463)
Buyer Log Wage		-0.0070*	-0.0080**
		(0.0031)	(0.0031)
Buyer Log Wealth		-0.0332**	-0.0313**
		(0.0048)	(0.0049)
Size of househ. buyer		-0.0387**	-0.0423**
		(0.0126)	(0.0127)
No. of children buyer		0.0320	0.0353
		(0.0205)	(0.0206)
High school buyer		-0.0021	-0.0051
		(0.0273)	(0.0274)
University buyer		-0.0106	0.0190
		(0.0289)	(0.0293)
Villa		-0.331**	-0.311**
		(0.0338)	(0.0347)
Flat		0.0592	0.0566
		(0.0338)	(0.0347)
Log square meters		-0.486**	-0.463**
		(0.0312)	(0.0398)
Month and year FE		yes	yes
Age dummy buyer		yes	yes
Municipality FE			yes
Observations	137,227	135,839	135,839

Notes: This table presents the results of a Cox proportional hazards model where the failure event is when a house is sold. Standard errors in parentheses.
** $p<0.01$, * $p<0.05$

Another probable consequence of a good housing match is that the owners are more likely to have children. People who want to have kids may put more weight on a better match, as moving is more costly with kids. To see whether this is reflected in my measure of housing quality, I measure the increase in the number of children under 18 in the household at end of year two after the year of housing purchase. I use a logit regression with increase in number of children as dependent variable, distance and other factors as regressors. I exclude households where the age of the buyer is 45 or above, as they are outside of the main childbearing age.

Table 6: Probability of children

New Children	(1)	(2)	(3)
Distance	**-0.0174****	**-0.0943****	**-0.0978****
	(0.0030)	(0.0037)	(0.0037)
Log price		0.792**	0.890**
		(0.0261)	(0.0473)
Buyer Log Wage		0.0330**	0.0293**
		(0.0043)	(0.0043)
Buyer Log Wealth		-0.0012	0.0043
		(0.0059)	(0.0061)
Size of househ. buyer		0.539**	0.546**
		(0.0179)	(0.0185)
No. of children buyer		0.749**	0.699**
		(0.0228)	(0.0236)
High school buyer		-0.114**	-0.144**
		(0.0325)	(0.0331)
University buyer		-0.0111	0.0270
		(0.0324)	(0.0335)
Villa			-0.0763*
			(0.0305)
Flat			-0.527**
			(0.0376)
Log square meters			0.0146
			(0.0466)
Age dummy buyer		yes	yes
Municipality FE			yes
Observations	99,263	99,207	98,073

Notes: This table presents the results of a logit regression where the outcome is an increase in the number of children under 18 two years after housing purchase. Only housing buyers below 45 years old are used.
Standard errors in parentheses.
** $p<0.01$, * $p<0.05$

Table 6 show that the probability of increasing the number of children in the household is decreasing in the distance between buyer and previous buyer (or increasing in the quality of the match).

The results presented in this section support the use of similarity between buyers as a proxy for match quality.

6 Conclusion

Housing search and matching models such as Anenberg and Bayer (2013) and Ngai and Tenreyro (2014) often explicitly or implicitly predict correlations between average match quality and easier observable housing market characteristics. A measure of match quality could be used to test these predictions. Knowing the match quality distribution of the population at a given time may also be helpful in predicting future levels of housing transactions.

In this paper, a housing search and matching model is used to show why similar buyers are more likely to be well matched, and can be predicted to pay more for their houses. In the model, successive owners of houses should share characteristics if the observable characteristics of a buyer are correlated with the buyer's preferences for housing.

I measure the similarity of a housing buyer and the previous buyer of the same house (who is now selling), and argue that this similarity can be used as a proxy for match quality. The similarity is measured as the Mahalanobis distance between characteristics of buyers and past buyers.

I utilize a rich set of data, 139,688 repeated housing sales, where I know the characteristics of the buyers in both set of transactions. Regressing prices on the similarity measure, I show that buyers who resemble previous buyers are paying more, also when a large number of observable characteristics are controlled for. This is in accordance with the model presented.

The distance measure is shown to be lower (similarity higher) in larger housing markets than in smaller markets, and negatively correlated with housing prices and transaction numbers. It can be seen as support for the thick-market effect in Ngai and Tenreyro (2014), where match quality and prices are higher when the stock of houses for sale is larger.

Supplementary analysis support that the distance between seller and buyer can be used as a proxy for match quality. Matches with low distance lead to slightly reduced hazard rate of reselling the house, and an increased probability of having children, both of which would be expected in a high quality match.

References

Anenberg, Elliot and Patrick Bayer (2013): “Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem”, NBER Working Paper 18980. National Bureau of Economic Research.

Bowlus, Audra J. (1995): “Matching Workers and Jobs: Cyclical Fluctuations in Match Quality”, Journal of Labor Economics, 13 (2): 335-350.

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

Centeno, Mário (2004): “The match quality gains from unemployment insurance”, Journal of Human Resources, 39 (3): 839-863.

Han, Lu and William C. Strange (2014): “Bidding Wars for Houses”, Real Estate Economics, 42 (1): 1-32.

Head, Allen, and Huw Lloyd-Ellis (2012): “Housing Liquidity, Mobility, and the Labour Market”, The Review of Economic Studies, 79 (4): 1559-1589.

Jovanovic, Boyan (1979): “Job matching and the theory of turnover”, The Journal of Political Economy, 87 (5): 972-990.

Kalmijn, Matthijs (1994): “Assortative Mating by Cultural and Economic Occupational Status”, American Journal of Sociology, 100 (2): 422–452.

Mahalanobis, Prasanta Chandra (1936): “On the generalized distance in statistics”, Proceedings of the National Institute of Sciences (Calcutta), 2: 49-55.

Mardia, Kantilal Varichand, John T. Kent and John M. Bibby (1989): Multivariate analysis, Academic Press.

Mare, Robert D. (1991): “Five decades of educational assortative mating”, American sociological review, 56 (1): 15-32.

Nenov, Plamen T., Erling Røed Larsen and Dag Einar Sommervoll (2016): “Thick-Market Effects, Housing Heterogeneity, and the Determinants of Transaction Seasonality”, The Economic Journal, forthcoming, doi: 10.1111/ecoj.12305.

Ngai, L. Rachel and Kevin D. Sheedy (2016): “The decision to move house and aggregate housing-market dynamics”, CFM Discussion Paper 2016-21.

Ngai, L. Rachel and Silvana Tenreyro (2014): “Hot and Cold Seasons in the Housing Market”, American Economic Review, 104 (12): 3991-4026.

Piazzesi, Monika, Martin Schneider and Johannes Stroebel (2015): “Segmented housing search”, NBER Working Paper 20823, National Bureau of Economic Research.

Pissarides, Christopher A. (1985): “Short-run equilibrium dynamics of unemployment, vacancies, and real wages”, The American Economic Review, 75 (4): 676-690.

Statistics Norway (2015): Income and wealth statistics for households, 2014, http://www.ssb.no/en/inntekt-og-forbruk/statistikker/ifhus/aar/2015-12-16.

Weiss, Yoram and Robert J. Willis (1997): “Match quality, new information, and marital dissolution”, Journal of Labor Economics, 15 (1): 293-329.

Appendix A: More on solving the model

To get to equation (6):

Combining the definition of η^*_{jk} in (5) and the owner value function (4) gives:

$\beta(B_j + V_k) + u = \frac{u+\gamma_{jk}+\eta^*_{jk}}{1-\beta(1-\delta)} + \frac{\delta V_k}{1-\beta(1-\delta)}$ which can be rewritten to (6)

Alternative for equation (10):

The steady state of value function of a buyer, B, (10), can alternatively be written:

$$B_j = \frac{1-\theta}{1-\beta}\sum_k(\frac{v_k}{v^t}\frac{h^*(\eta^*_{jk})}{1-\beta(1-\delta)})$$

Equation (10) equals equation (7) in Ngai and Tenreyro (2014).

To get to equation (17):

$E[P_{jk}] = (1-\theta)\frac{u}{(1-\beta)} + \theta E[H(b_j, v_k, \eta)|\eta > \eta^*_{jk}]$

or $(1-\theta)\frac{u}{(1-\beta)} + \theta(H(b_j, v_k, \eta^*_{jk}) + E[S_{jk}|\eta > \eta^*_{jk}])$

We know that $E(S_{jk}|\eta > \eta^*_{jk}) = \frac{E[\eta-\eta^*_{jk}|\eta>\eta^*_{jk}]}{1-\beta(1-\delta)}$

$$\begin{aligned} H(b_j, v_k, \eta^*_{jk}) &= \beta(B_j + V_k) + u = u + \beta(\left(\frac{1-\theta}{(1-\beta)(1-\beta(1-\delta))}\right)(\sum_k[\frac{v_k}{v^t}h^*(\eta^*_{jk})]) \\ &+\frac{u}{1-\beta} + (\frac{\theta}{(1-\beta)(1-\beta(1-\delta))})(\sum_j[\frac{b_j}{b^t}h^*(\eta^*_{jk})]) \\ &= \frac{u}{1-\beta} + \beta(\left(\frac{1-\theta}{(1-\beta)(1-\beta(1-\delta))}\right)(\sum_k[\frac{v_k}{v^t}h^*(\eta^*_{jk})]) + (\frac{\theta}{(1-\beta)(1-\beta(1-\delta))})(\sum_j[\frac{b_j}{b^t}h^*(\eta^*_{jk})]) \end{aligned}$$

So $E[P_{jk}] = \frac{u}{(1-\beta)} + \theta[\beta(\left(\frac{1-\theta}{(1-\beta)(1-\beta(1-\delta))}\right)(\sum_k[\frac{v_k}{v^t}h^*(\eta^*_{jk})]) + (\frac{\theta}{(1-\beta)(1-\beta(1-\delta))})(\sum_j[\frac{b_j}{b^t}h^*(\eta^*_{jk})]) + \frac{E[\eta-\eta^*_{jk}|\eta>\eta^*_{jk}]}{1-\beta(1-\delta)}]$, which can be rewritten as (17).

Appendix B: Extra figures and tables

Figure A.1: Mahalanobis distance and transactions

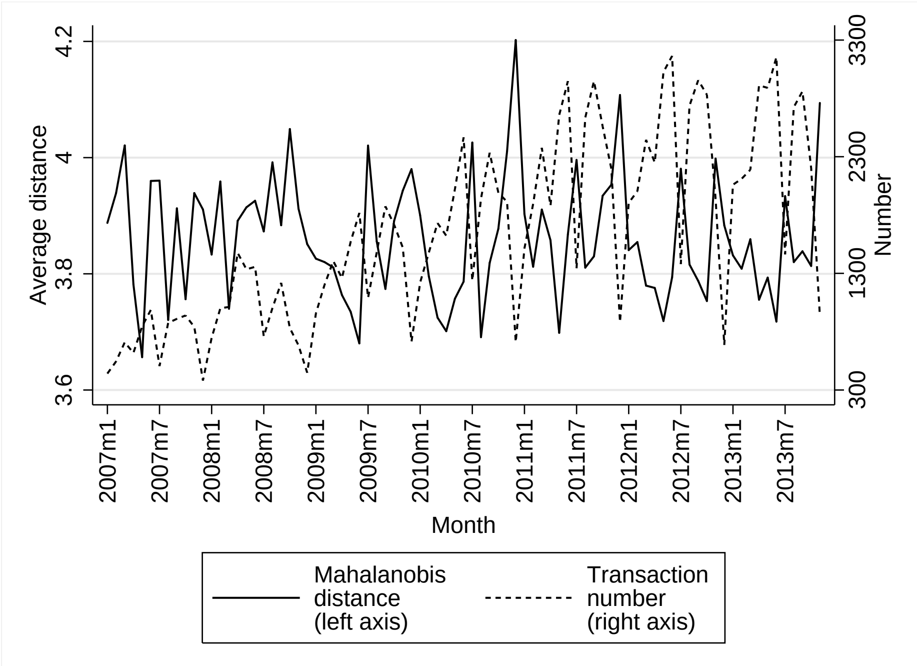


Figure A.1 plots the average Mahalanobis distance and housing transaction number over time.

Table A.1: Results, row-houses

Log price	(1)	(2)	(3)	(4)	(5)
Distance	**-0.0214****	**-0.0028****	**-0.0028****	**-0.0020****	**-0.0013****
	(0.0011)	(0.0004)	(0.0004)	(0.0003)	(0.0004)
Log appraisal value		0.979**	0.972**	0.829**	0.822**
		(0.0023)	(0.0023)	(0.0048)	(0.0048)
Buyer log Wage				0.0013**	0.0013**
				(0.0003)	(0.0003)
Buyer log Wealth				0.0039**	0.0038**
				(0.0005)	(0.0005)
Size of househ. buyer				-0.0004*	-0.0012
				(0.0014)	(0.0014)
No. of children buyer				-0.0012	-0.0005
				(0.0019)	(0.0019)
High school buyer				0.0050	0.0057*
				(0.0027)	(0.0027)
University buyer				0.0209**	0.0207**
				(0.0027)	(0.0027)
Log square meters				0.0792**	0.0829**
				(0.0045)	(0.0046)
Size of househ. seller					0.0023
					(0.0016)
No. of children seller					-0.0015
					(0.0020)
High school seller					0.0109**
					(0.0026)
University seller					0.0203**
					(0.0027)
Month and year FE			yes	yes	yes
Age dummy buyer				yes	yes
Municipality FE				yes	yes
Age dummy seller					yes
Observations	21,123	11,668	11,668	11,627	11,596
R-squared	0.019	0.941	0.943	0.959	0.960

Standard errors in parentheses.

** p<0.01, * p<0.05

Table A.2: Results, villas

Log price	(1)	(2)	(3)	(4)	(5)
Distance	**-0.0133****	**-0.0009****	**-0.0009****	**-0.0007****	**-0.0003**
	(0.0009)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Log appraisal value		0.994**	0.992**	0.923**	0.918**
		(0.0013)	(0.0013)	(0.0026)	(0.0026)
Buyer log Wage				0.0007**	0.0007**
				(0.0002)	(0.0002)
Buyer log Wealth				0.0040**	0.0040**
				(0.0004)	(0.0004)
Size of househ. buyer				0.0016	0.0016
				(0.0010)	(0.0010)
No. of children buyer				-0.0008	-0.0009
				(0.0014)	(0.0014)
High school buyer				0.0096**	0.0092**
				(0.0018)	(0.0018)
University buyer				0.0224**	0.0215**
				(0.0020)	(0.0020)
Log square meters				0.0124**	0.0141**
				(0.0027)	(0.0028)
Size of househ. seller					0.0011
					(0.0010)
No. of children seller					-0.0008
					(0.0015)
High school seller					0.0115**
					(0.0018)
University seller					0.0187**
					(0.0020)
Month and year FE			yes	yes	yes
Age dummy buyer				yes	yes
Municipality FE				yes	yes
Age dummy seller					yes
Observations	52,776	27,748	27,748	27,602	27,491
R-squared	0.005	0.956	0.957	0.964	0.964

Standard errors in parentheses.

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

Table A.3: Results, flats

Log price	(1)	(2)	(3)	(4)	(5)
Distance	**-0.0231****	**-0.0026****	**-0.0029****	**-0.0019****	**-0.0010****
	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0003)
Log appraisal value		0.913**	0.901**	0.675**	0.669**
		(0.0016)	(0.0016)	(0.0026)	(0.0027)
Buyer log Wage				0.0000	0.0001
				(0.0002)	(0.0002)
Buyer log Wealth				0.0065**	0.0064**
				(0.0003)	(0.0003)
Size of househ. buyer				0.0011	0.0006
				(0.0008)	(0.0008)
No. of children buyer				-0.0107**	-0.0104**
				(0.0014)	(0.0014)
High school buyer				0.0115**	0.0121**
				(0.0019)	(0.0019)
University buyer				0.0273**	0.0272**
				(0.0020)	(0.0020)
Log square meters				0.219**	0.225**
				(0.0026)	(0.0027)
Size of househ. seller					0.0026*
					(0.0011)
No. of children seller					-0.0076**
					(0.0016)
High school seller					0.0154**
					(0.0020)
University seller					0.0299**
					(0.0020)
Month and year FE			yes	yes	yes
Age dummy buyer				yes	yes
Municipality FE				yes	yes
Age dummy seller					yes
Observations	63,350	38,771	38,771	38,627	38,460
R-squared	0.023	0.894	0.899	0.922	0.923

Standard errors in parentheses.
** p<0.01, * p<0.05

Table A.1-A.3 show the specifications from Table 3 run separately on the sample of row-houses, villas and flats. The size and significance of the distance measures roughly holds for all three specifications.

Table A.4: Ownership length and distance

Log price	(1)	(2)
Distance	**-0.00139****	**-0.00047**
	(0.00017)	(0.00032)
Ownership length		0.00019**
		0.00003
Distance x ownership length		-0.00002**
		(0.00001)
Log appraisal value	0.801**	0.801**
	(0.0018)	(0.0018)
Buyer log Wage	0.00066**	0.00066**
	(0.00013)	(0.00013)
Buyer log Wealth	0.00506**	0.00507**
	(0.00023)	(0.00023)
Size of househ. buyer	0.0013*	0.0013*
	(0.0006)	(0.0006)
No. of children buyer	-0.0027**	-0.0027**
	(0.0009)	(0.0009)
High school buyer	0.0098**	0.0098**
	(0.0013)	(0.0013)
University buyer	0.0255**	0.0254**
	(0.0013)	(0.0013)
Villa	-0.0057**	-0.0054**
	(0.0014)	(0.0014)
Flat	-0.0034*	-0.0033*
	(0.0015)	(0.0015)
Log square meters	0.113**	0.113**
	(0.0018)	(0.0018)
Month and year FE	yes	yes
Age dummy buyer	yes	yes
Municipality FE	yes	yes
Observations	77,687	77,687
R-squared	0.942	0.942

Standard errors in parentheses.

** p<0.01, * p<0.05

Table A.4 shows, in the first column, the results from column (4) in Table 3. In the second column, I have included ownership length in months as a regressor, and an interaction term between ownership length and Mahalanobis distance. The negative, significant coefficient on the interaction term indicates that the Mahalanobis distance has a larger impact on price with longer ownership length of the seller. The coefficients not affected by the interaction term are not noticeably changed by its inclusion.

Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISSN: 1892-753X

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist