Aalto University
School of Business

Analysis of expected sale time and behavior on the Finnish housing market

Finance

Master's thesis

Pontus Oinonen

2013

Department of Finance
Aalto University
School of Business

Aalto University
School of Business

Analysis of expected sale time and behavior on the Finnish housing market

Master's Thesis
Pontus Oinonen
Spring 2013
Finance

Approved in the Department of Finance __ / __20__ and awarded the grade

____

Author Pontus Oinonen

Title of thesis Analysis of expected sale time and behavior on the Finnish housing market

Degree MSc

Degree programme Finance

Thesis advisor Matti Keloharju

Year of approval 2013 **Number of pages** 73 **Language** English

PURPOSE OF THE STUDY

The purpose of this thesis is to determine how expected sale time of housing in Finland is formed and which factors have an effect on it. I use marketing time of dwellings as a proxy of expected sale time. This is a descriptive study in nature, because no earlier papers with Finnish data exist to compare the results with. The analysis covers housing features as well as its location. In addition, I study macro-economy and date of market entry related variables' effects on expected sale time. For most individuals, housing transactions are largest investment decisions of their lives and I aim to explain the high variation in the sale times within the Finnish housing market and in comparison with other asset classes.

DATA

I study my research questions mainly based on housing advertisements from January 2004 to July 2012. The advertisement data was collected from three sources: Etuovi, Oikotie and Kiinteistömaailma. In addition, I collected real transaction data from asuntojen.hintatiedot.fi and various background data from Statistics Finland.

RESULTS

The results show that nearly all available housing and listing date related features have a statistically significant relationship with expected sale time. I also find that submarkets within certain cities do not vary much from the city consensus and more variation is present at times with poor state of economy. Results suggest that a seller with no rush to sell should wait for the optimal listing time to ensure her dwelling to raise interest in the market. I find that buyers' search criteria selection has a clear link to expected sale time as well. With a careful selection of listing price the seller may lower her dwelling's expected sale time.

Keywords Housing market, time-on-market, submarkets, market characteristics, behavior

Tekijä Pontus Oinonen

Otsikko Analyysi odotetusta myyntiajasta ja käyttäytymisestä Suomen asuntomarkkinoilla

Tutkinto KTM

Tutkinto-ohjelma Rahoitus

Ohjaaja Matti Keloharju

Hyväksymisvuosi 2013 **Sivumäärä** 73 **Kieli** Englanti

TUTKIELMAN TAVOITTEET

Tutkielman tavoitteena on määritellä mistä asunnon odotettu myyntiaika koostuu ja mitkä tekijät tähän Suomessa vaikuttavat. Käytän asuntojen markkinointiaikaa odotetun myyntiajan ennustajana. Tutkielma on luonteeltaan kuvaileva, koska vastaavaa tutkimusta vertailukohdaksi ei ole Suomessa ennen tehty. Analyysi kattaa asunnon ominaisuudet ja sijainnin. Tämän lisäksi tutkin kansantalouden makrokuvaajien sekä asunnon listausajankohdan vaikutusta odotettuun myyntiaikaan. Useimmille ihmisille asuntokauppa on elämän suurin investointipäätös ja tavoitteenani on selittää asuntojen myyntiaikojen suurta vaihtelua Suomen asuntomarkkinoiden sisällä ja suhteessa muihin sijoitusinstrumentteihin.

LÄHDEAINEISTO

Tutkimuskysymyksiini vastaan pääasiallisesti asuntojen myynti-ilmoituksiin aikavälillä tammikuu 2004 ja heinäkuu 2012 perustuen. Ilmoitukset on kerätty kolmesta lähteestä: Etuovelta, Oikotieltä ja Kiinteistömaailmalta. Tämän lisäksi olen kerännyt tietoa toteutuneista kaupoista asuntojen.hintatiedot.fi sivustolta sekä monenlaista taustatietoa Tilastokeskukselta.

TULOKSET

Tulokset osoittavat, että lähes jokaisella saatavissa olevalla asuntoon tai listauspäivään liittyvällä tekijällä on tilastollisesti merkittävä suhde odotettuun myyntiaikaan. Totean samalla, etteivät osamarkkinat tutkimani kaupunkien osalta tuo suurta vaihtelua tuloksiin suhteessa koko kaupunkiin kokonaisuutena. Enemmän vaihtelua kuitenkin esiintyy heikon talouden tilan aikaan. Tulokset indikoivat myös, että myyjän, jolla ei ole kiire, tulisi odottaa optimaalista aikaa osallistua markkinoille varmistaakseen kiinnostuksen omaan asuntoonsa. Löydän myös selvän yhteyden ostajien hakukriteerien ja odotetun myyntiajan välillä. Tämä osoittaa, että myyjä voi vaikuttaa asuntonsa odotettuun myyntiaikaan muun muassa tarkasti harkitulla listaushinnalla.

Avainsanat Asuntomarkkinat, markkinointiaika, osamarkkinat, markkinoiden ominaispiirteet, käytös

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

For most people, housing transactions are the most significant investment decisions of their lives. Unprepared individuals may get surprised about the transaction time the process takes. Other people are aware of this and take the longer transaction time as granted. In this thesis I seek patterns and study market characteristics to minimize the information asymmetries surrounding the time it takes to trade a certain dwelling. My main objective is not to find how to truncate the transaction time, but I seek to explain why house A will be sold earlier than house B or the other way around. Most of financial asset research relates to securities market, which is considered more appealing and where data is easier to collect, but popularity of real estate research has seemingly increased in past decades. I will study one of the housing markets' important separating characteristics –liquidity.

In housing, liquidity can be considered as sale time and transaction costs. When selling few liquid stocks in a stock exchange at market price, one can be rather sure to get them sold instantly and with a relative small cost. However, housing markets seem to be less liquid as explained later. A housing transaction, even at a market price, can take a long time before completion. Online market places have reduced the search time remarkably (see e.g. Bakos 1997), but housing features, such as heterogeneity, still reduce the probability of a quick transaction. In addition to time consumption of the trade, the physical costs of selling a house may get remarkable high. In a typical setting, buyers search for houses with desired attributes in the market and sellers receive offers from them. This process requires time and I'm interested whether I can find systematic differences in marketing time with different housing or market features.

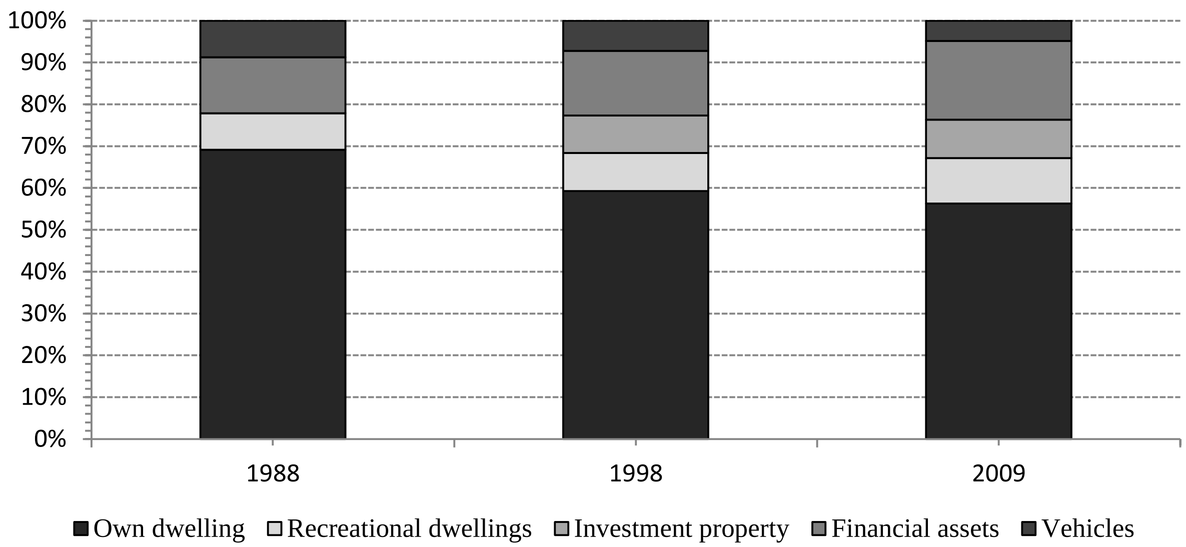
My focus is on dwellings of all kind in Finland excluding other type of real estate as business facilities and summer cottages. I have personal interest towards real estate as an investment vehicle and consider myself a future real estate investor. In addition, I want to contribute to existing literature with a Finnish perspective and thus do my part in improving the grounding for future research.

I follow the logic of known real estate liquidity studies (see e.g. Anglin et al. (2003) or Pryce and Gibb (2006)) and combine other approved methods such as behavioral studies on top of them to have a diverse research. Finnish studies on real estate are mainly about consumer preferences (see e.g. Juntto, 2007 or Tyvimaa and Kananen 2011), pricing (see e.g. Oikarinen, 2007) or structural development (see e.g. Huovari et al. 2002 and 2006). What makes this study even more unique is that most of papers published in Finland are done by research institutions such as Statistics Finland, Pellervo Economic Research Institute or KTI Kiinteistötieto Oy. A great challenge with this topic is the lack of popularity among academics leading to low number of recognized theories. In addition, most of newer papers disprove methods used in older papers narrowing the possible testable theories even further. To my best knowledge there is no other paper studying marketing times with Finnish data available. This motivates me to study the Finnish housing markets to search for personalities and consistent patterns for expected sale time estimation.

Dwellings people live in make up more than the half of the Finns' gross wealth. If I add investment property and recreational dwellings to it, I get a total contribution of 76 percent for real estate in 2009, as shown by Figure 1. This is why my study should be interesting and helpful for many people outside academics as well.

Figure 1 Distribution of Finnish households' gross assets

This column chart shows how the gross wealth of an average Finnish household's was distributed in the end of last three decades as percentages of total gross assets. Year 1988 did not include separate data of investment property and it wasn't necessarily zero. Data is from the Statistics Finland.



If I take the investor's point of view, understanding features of different assets is a benefit when constructing one's own investment portfolio. Real estate can be used as part of portfolio whereas other assets. Motivations to hold real estate include risk decreasing and positive cash flow securing in terms of rents as discussed later in this paper. Friedman (1971) continued the work of Markowitz (1952) to apply real estate in modern portfolio theory. He wrote that real estate works just as stocks when constructing a portfolio. Friedman states that real estate can provide higher risk adjusted returns as stocks when the target's risk and return are known for calculating this. In addition, real estate returns have a positive correlation with inflation allowing real estate investments to serve as an inflation protection (see e.g. Firstenberg et al. (1988)).

Despite the important investor point of view discussed above, one must remember that real estate, especially housing, is often not acquired for investment purposes, but for example as a place to live in. This could be one factor causing housing markets to be less liquid than those of public stocks. Public stocks can have strategic owners as well (as the state in Finland), but usually their free float is large enough to enable more liquid markets than with housing.

1.2 Marketing time definition and potential limitations

In this paper I study marketing time of houses. In this chapter I explain what I mean with marketing time and the differences to similar measures. One commonly discussed measure is the sale time of houses. The difference between these two concepts is that whereas sale time is the time between the listing of a house by the seller until the money is transferred by the buyer (Belkin et al. 1976), marketing time does not necessarily confirm a sale, only that the marketing of the property has ended. This means that with high probability there are observations that have never lead to a transaction. These are called censored observations as I later discuss. Realtors and market places only know the time they have advertised the house, thus the marketing time. The main data used in this study bases on marketing time and therefore uncertainty of transactions exists. Most papers on transaction time use term Time-On-Market (TOM) regardless of their data type (sale time or marketing time) thus making me feel comfortable using an own definition of TOM as well. The reason for the vague use of this term is that the whole selling time is not recorded by any actor. Only the seller might know the actual time when she starts the selling process and when the transaction occurs. Therefore, even if I had the transaction dates available for all observations, it would still be a proxy of the sale time.

One thing to consider is that individual sellers may have tried to sell the house by themselves before contacting a realtor. Also, the seller may have changed the realtor during the selling process or delisted her target for a period during the selling process. The latter actions may increase interest in the house as discussed later on. These phenomena will cause outliers in the data that I’m unable to fully distinguish. I believe marketing times I use will serve a good proxy and quality of results will not suffer from this method.

1.3 Research questions

This paper is first of its kind in Finland, which makes me eager to have multiple research questions allowing me to study the subject widely while covering multiple issues. The key research questions are:

1. How does marketing time of a dwelling depend on housing features, location and local macro-economic factors?
2. Do submarkets exist and how they affect the marketing time in various Finnish housing markets?
3. Are there behavioral patterns decreasing or altering liquidity on the Finnish housing market?

1.4 Contribution to existing literature

I will use tests performed in other countries to serve insight to Finnish housing markets. No one has done exactly the same kind of study and I hope my results will be interesting to read and that they help the reader to identify the Finnish housing market or differentiate them from others. In addition, I test for behavioral patterns mainly studied with asset returns to see if they are applicable for marketing time as well.

My results suggest that most housing, location and market entry date related variables have significant impact on expected sale time. I also find that submarkets within certain cities do not vary much from the city consensus and more variation is present at times with poor state of economy. In addition, a seller with no rush to sell should wait for the optimal listing time to ensure her dwelling to raise interest in the market. I find that buyers’ search criteria selection has a clear link to expected sale time as well. With a careful selection of listing price the seller may lower her dwelling’s expected sale time.

1.5 Structure of thesis

The rest of my thesis is organized as follows: First I explain groundings of the work helping reader to understand the business field the paper is about. Secondly I conclude theoretical aspects and previous literature on housing markets and other key elements of this study. Then I introduce the methodology and the data I use in this study. Lastly I present the results and summarize the thesis with conclusions and suggestions for further research.

2 PRACTICAL SETTING

This part is for the reader to understand how the Finnish housing market has developed to its current state and the basic characteristics of the market. With housing markets, it is important to have a long period of time to see structural changes as they are slower than those of asset markets of higher liquidity.

2.1 DEVELOPMENT AND STRUCTURE OF THE FINNISH HOUSING MARKET

The Finnish housing market has developed a lot in past decades. This is largely due to deregulation of the markets and international influences. The volatility of Finnish housing market started when international shocks begun to affect local economy. This was in the 1970s although the post-oil crisis drop in real prices was due to high inflation. In 1986 the gradual opening of capital markets with good employment situation and stable economic growth phase lead to quick increase in dwelling prices as people got their loans easier (Huovari et al. 2006). Same time the down payment ratios lowered and this led to huge growth of credit and a housing market boom (Oikarinen 2007). This and the later depression of early 1990's were reasons for further development.

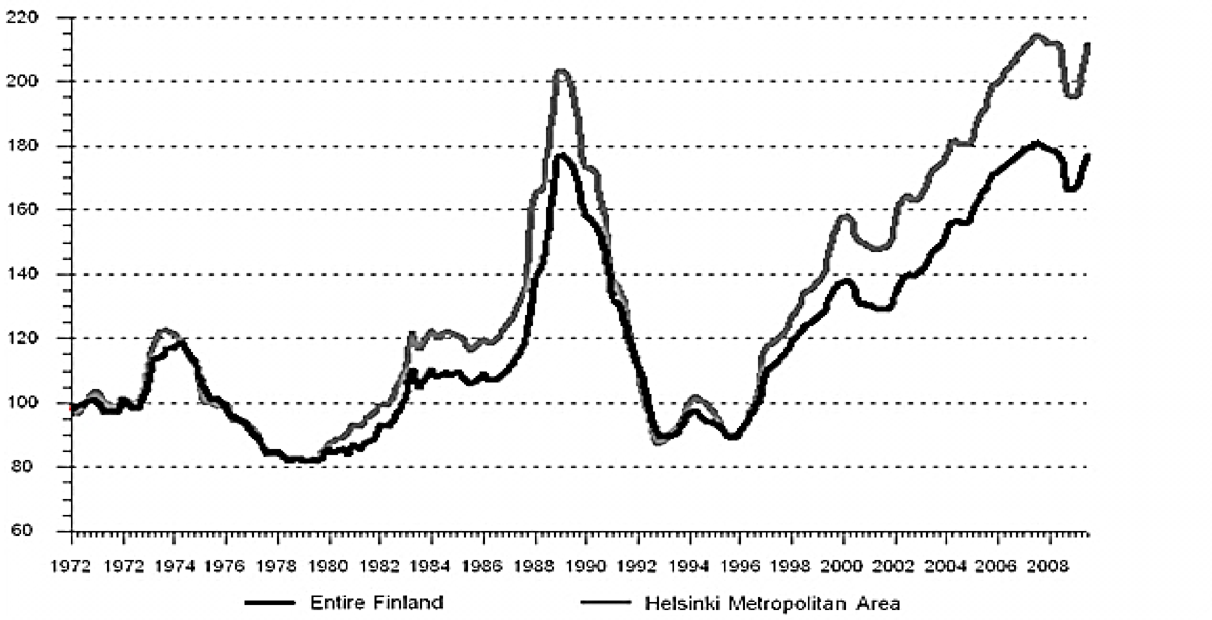
Laws on tax deductions on mortgage and interest payments have varied in fundamentals and affected willingness to house purchases. What is interesting from a housing investor's point of view, the rental markets were not fully opened until 1995, after a three year gradual development. Only since then the investors have had the freedom to decide for rents (except of increases in current contracts). A large portion of houses in Finland is publicly regulated and only those under private control face no regulation of pricing. (Oikarinen 2007)

Figure 2 shows how the real prices of housing have developed during the past four decades. As discussed earlier, the financial markets were gradually opened in mid-1980 and the steep upward curve is the result of overheated housing markets. The following drop was partly due to the overheated markets, but highly affected by the general recession in Finland.

Figure 2 also shows that Helsinki metropolitan area (HMA) prices in 1989 peaked higher than Finland as a whole. The same phenomenon has occurred with every time the prices have increased in the past 40 years. On the other hand, HMA has traditionally survived the price drops quicker than Finland in general and has never sunk below the country index. My data includes city of Helsinki and it is important to understand the heavy weight this area represents when analyzing Finnish data.

Figure 2 Real price index for old Finnish apartment buildings

This figure illustrates real price development from Q1/1970 – Q2/2009. The base year is set at 1970 with the value of 100. Finland as a whole and Helsinki metropolitan area are shown on own lines. Source of data is Statistics Finland.

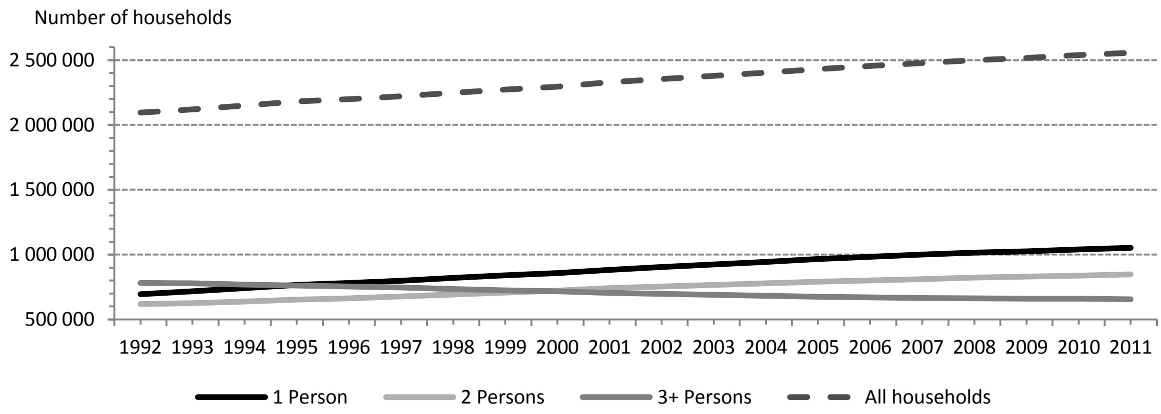


We are currently again at the price levels of the previous housing market bubble and the discussion is lively whether another burst will occur. KTI (2012) reports that the Finnish housing market currently faces mild uncertainty in the short run, but whether the movement is up or down it will not be a large one. They remind that from investors' point of view a temporary drop in housing prices is not crucial, because at this time popularity of rented houses tends to rise. If the investor can renegotiate rents, this can level the situation. KTI also points out that investments in housing should be considered as long term assets.

Figure 3 illustrates how the number of different size households has developed in 1992-2011. It is fairly clear that larger households have diminished and single and double households have increased in popularity steadily over time. This results from demographic changes in Finland. Number of people living alone or two by themselves has grown with traditional family model diminishing in the society. When the number of average household members lowers, together with population growth, it increases the number of households as Figure 3 suggests. A rational conclusion of this would be that new apartment houses would be smaller in average to serve smaller households. The logic here is that number of smaller households would rise especially in urban areas and mostly apartment buildings are constructed in them. However, Figure 4 doesn't support this suggestion; average new apartment size has grown from 75 to 82 square meters in 1990 to 2011. On the other hand, this information supports the suggestion of increased living standards in Finland.

Figure 3 Development of household unit size in Finland 1992-2011

This figure shows how the household composition has developed 1992-2011 in Finland. Number of persons indicates how many persons live in the same household, thus in the same dwelling. The data is from Statistics Finland.



Statistics Finland report that the number of households living in apartment houses and detached houses has increased in the past two decades whereas row houses and similar have remained their popularity constant. Detached houses and similar contributed for 89 percent of all housing buildings in the end of 2011 in number terms. This relation is clearly explained by that apartment houses shelter many more households than the detached houses.

Figure 4 shows the amount of new construction permits in Finland 1990-2011. It indicates that relation of permitted floor space between different housing types has remained quite the same except of a high season of detached houses in mid-2000. All three economy dips during the period, the current, the internet bubble and the early 1990's, are clearly visible in this figure as lower demand for housing construction. Construction permits on all housing types seem to reflect to economic changes. Permits on detached houses have the highest volatility which can be explained by the relatively high amount of them financed in individuals. During an economic downturn the state may start constructions of apartment buildings as part of their reflationary policies and individuals are more vulnerable.

Figure 4 Construction permits and completed housing

This figure shows the number and area of new construction permits 1990-2011. The scale on the left-hand side shows amount of permitted area in square meters plotted by the lines. Columns indicate the number of permits shown in the scale on the right-hand side. The data source is Statistics Finland.

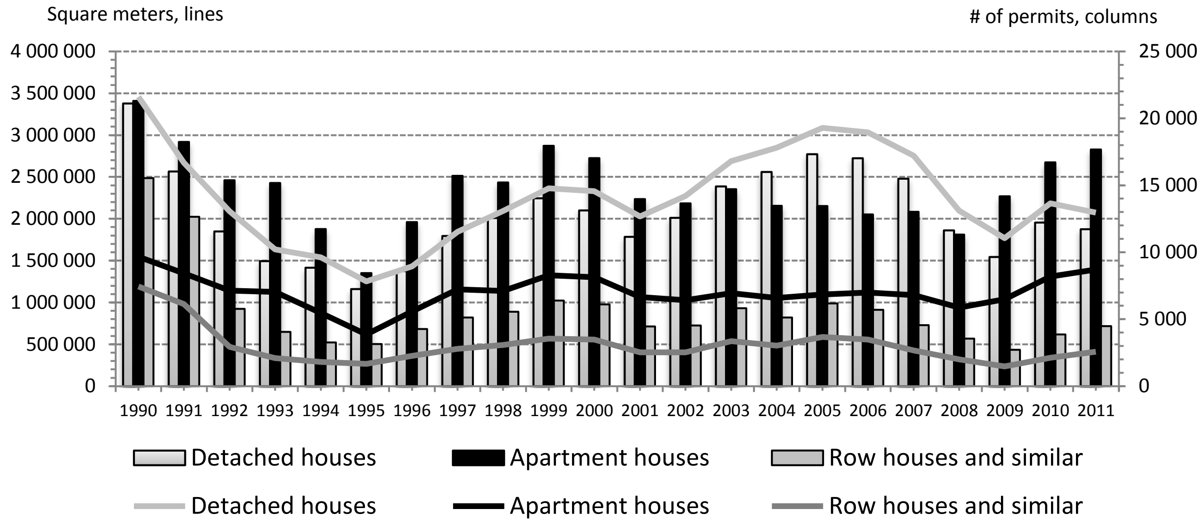
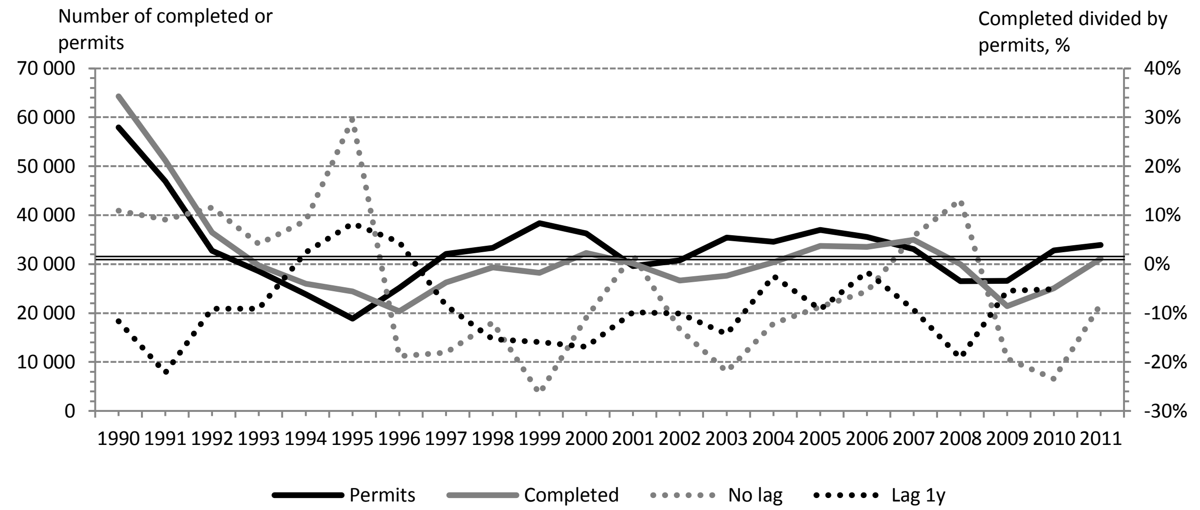


Figure 5 Relation of construction permits and completed housing

In this figure we see how the relation of construction permits and completed housing has developed 1990-2011. Primary vertical axis represents amounts and the secondary axis ratio of completed housing divided by initial permits. Dotted lines show the same relation, but with and without a lag in construction. Data source is Statistics Finland.



In Figure 5 I show how the number of construction permits correlates with completed housing. It is clear that construction takes time and therefore completed houses lag behind permits. However, Figure 5 serves interesting information. Prior the internet bubble, mainly in 1999 and 2000, the number of permits was very high, but the number of completed housing in following years didn’t get near the permit level. This shows that many plans were left unrealized as the economy turned bad. The two lines for permits and completed buildings cross every time the economy turns with a small lag. Similar calculation can be done to all years and see how the rate of permits and completed housing fluctuates. Dotted lines show this contribution in Figure 5. We see that without a lag the relation fluctuates heavily, because construction takes time and when the economy turns bad not all building sites are closed and completed buildings still arise. When I analyze the same relation with a one year lag in number of permits, e.g. completed in 2000 divided by permits in 1999, the line shows a proxy of completion rate of buildings. It is clear that the number of completed buildings is lower than number of permits if we don't account for illegal construction.

2.2 Consumption preferences

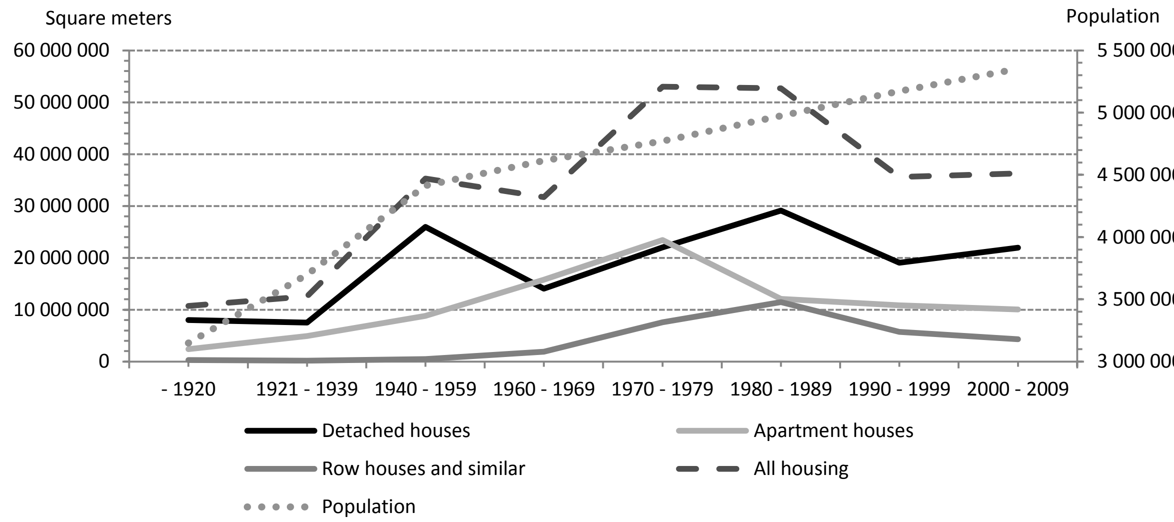
Overheating of markets in late 1980 also enhanced perhaps the most discussed development in Finnish housing markets of diverging of areas in terms of population also known as urbanization. One example of this is people from small towns going to a university or a job in a bigger city never to return. Another cause of leaving is the industrial cutbacks. There are lots of Finnish towns build around a factory or an industrial function. In a case of a shutdown people often need to move elsewhere for a job. Non-returners leave empty dwellings to these small towns thus lowering the capital value of other home owners, i.e. supply grows while demand lowers (Huovari et al. 2002). Social effects may rise in importance as well if for example fewer children are born and the local school therefore shut down.

They continue by arguing that larger cities, especially Helsinki, face some challenges more powerful than other cities. Infrastructural development is hard to keep up with the increasing population. Increasing population also reflects the housing markets with higher rents and dwelling prices due to the higher demand. This effect can be explained with the four quadrant model I introduce in chapter 3.2. When the demand grows and new housing construction takes time, the rational solution is for the prices to go up as the equation of supply and demand says. This phenomenon can harm growing companies who have trouble getting new labor force in the city. Winners in this contest are the cities that offer jobs and education institutions. Municipalities around these growing cities have witnessed increased population as well. These surrounding areas are alternative living spaces for people working in a city, but unsatisfied with the high living costs. Development of public transportation shortens the distance and allows for people to live outside of cities as well. Urbanization rate in Finland was at 68 percent level in 2011[1].

Urbanization in Finland started relatively late, in the 1950's after the Second World War. The largest migrations from rural areas into cities took place in the 1960's and 1970's. To serve this inflow of people a number of cities started massive constructions of concrete suburbs. Most of these suburbs consisted of apartment houses that worked in form of limited companies (Viitanen et al. 2003). This construction peak of apartment houses is easily observed from Figure 6. The rapid construction focused on quantity of housing rather than quality as they continue. Currently many of these concrete blocks are quickly deteriorating and need either a great repair or demolition. During the 1980's the focus on construction shifted into quality and improving living standards. Figure 6 shows that row houses seemed a trend for few decades and have lately been less popular. Most of housing stock in Finland is built in the 1970's and 1980's. The economic state explains part of variation in construction behavior as explained with Figure 5, but other factors such as population growth have impact as Figure 6 suggests.

Figure 6 Construction decades of current housing buildings by area

This figure shows the construction decades of current housing buildings by building type and population in Finland. Values on the left-hand scale represent square meters of living space and the other scale is for population. The unsymmetrical time line comes with data limitations. With a stable time line the growth of construction from early on would be more visible as it is now. Source of data is Statistics Finland.



Juntto (2007) reports survey results that 86 percent of Finnish people want to own a house while 63 percent actually own one. Thus one could say that Finnish people appreciate the

[1] http://www.findikaattori.fi/fi/56

value of owning a house. Another finding of this survey is that respondents on average want larger houses. Statistics Finland tells us that people have been able to follow this will. They report a steady increase in average house size from 60 square meters in 1970 to 79.8 square meters in 2011. However, this increase could also be explained by increased living standards of the Finns. A very high portion of the Finns (84 percent) are either very or quite satisfied with their current dwelling. Despite this result only 50 percent feel they are in dwelling where they target to live in.

Another survey conducted by Tyvimaa & Kananen (2011) also reports that Finns are in general satisfied with their current housing. However, 47 percent of respondents answered positive when asked if they are considering moving in the following 12 months. Need for extra space is ranked the top reasoning for moving in this survey as well. The most important factor when looking for a new home was that the area surrounding is safe and secondly that the new home would feel safe to live in. This finding is also in line with the survey by Juntto (2007).

The search process has also developed during the last couple of decades. This is mainly due to lowered search costs provided by online market places. Here, sellers may add photos and details of their commodity with contact information and possibility for dynamic pricing. Internet has been the main source of information and realtor finding for sellers already for many years in Finland (Uusitalo2008). He states that before the internet, the main marketing channel was the printed advertisements in newspapers. The shifting of consumption preferences has also lowered market entrance costs and we have witnessed a growth of new real estate agencies, many of which act in the internet only. We have also seen a formation of online market places that only post sellers’ advertisements and let buyers to contact the sellers directly without a realtor in between.

3 THEORETICAL BACKGROUND

The main characteristic of dwellings under interest in this paper is liquidity. Morawski (2008) lists 28 different definitions for the word liquidity collected from various online lexicons. When it comes to real estate, I find the definition of Wood & Wood (1985) a good one. They define real estate liquidity as “the inverse of the amount of time that elapses between the decision to sell a security and the receipt of the full market value by the seller.”

Liquidity can be perceived as asset or market liquidity. Asset liquidity refers to time and costs required to liquidate a certain investment. There is a common agreement that many other financial assets like stocks are more liquid than any direct real estate investments. This derives from real estate characteristics such as heterogeneity and spatial fixity as discussed in the next chapter.

What comes to market liquidity, real estate markets have features that prevent them to be as liquid as stock markets. Especially less populated areas, such as most of Finland, have fewer buyers and sellers acting at the same time. Even fewer are looking for a dwelling with exactly the features provided and therefore the supply and demand do not always meet. High transaction costs, segmentation and information asymmetries among other things delay transactions and lowers the market liquidity as discussed in the coming chapters. These principles go for real estate markets in general and can be applied to housing markets as well.

3.1 DIFFERENCING CHARACTERS OF DWELLINGS AS INVESTMENT VEHICLES

Like with other assets, houses should be priced with forecasted future cash flows (Oikarinen 2007). However, buying a house differs a lot from buying another type of asset class. Firstly, people buy houses to have a shelter, a place to live. It doesn’t matter if it’s owned or a rental place, but they need to have one. This is an important feature that decouples dwellings from other investments. One notable problem with housing is that the seller has more information of the target than the buyer and the buyer may be unwilling to disclose the price she would pay for the dwelling (Kramer 1999)

Dwellings are commodities as any, but the unique features discussed above and later on in this chapter explain why housing markets work quite differently from other commodity markets. Arnott (1987) lists many of these characteristics such as durability, spatial fixity, invisibility, complexity and non-convexities in production. Next I present features that crucially affect liquidity of dwellings in more detail.

3.1.1 High transaction costs

Transaction costs of houses derive from direct costs such as realtor fees and indirect costs like sale time as explained previously. Taloussanomat, a leading Finnish internet based business newspaper studied in 2009 the realtor fees for a €280 000 two-room apartment in central Helsinki. Results varied between circa 1-5 percent of purchase price (€2 440-€14 550).[2] In Finland the sales profit is under same capital tax rate as stocks if the seller hasn't lived in the house for two years. At this point the sales profit becomes non-taxable profit. A recent survey by Helsingin Sanomat, a major Finnish newspaper, shows the average commission rates for seven largest real estate agencies in Finland. They range from 2.42 to 3.66 percent from the transaction price. However, all agencies agreed that price should not be the only factor to consider when choosing an agency. Larger agencies tend to be more expensive, but offer wider networks and larger marketing machineries for their commissions.[3]

According to e.g. Morawski (2008) trading activity should be weaker in markets with higher transaction costs. The higher the transaction costs, the longer marketing times should be observed. This is because sellers are reluctant to trade houses often if they have to face a large transaction cost each time. Yang & Yavas (1994), Frew (1985) and Jud et al. (1996) among others investigate the impact of real estate brokerage firms on market. They study e.g. if some brokers are quicker to sell or able to close a higher price. Vast majority of my data refers to transactions via real estate brokers, but I will not rank agencies as I have no data on them.

One possible way of mitigating transaction costs is to sell a dwelling without an agent and save the realtor fee. This requires knowledge on valuating the dwelling as well as on the documents required and perhaps some eye for decorating a dwelling to attract potential buyers. The person would have to do marketing by herself. All of these are services you normally pay for the realtor to do. Direct costs of selling a dwelling alone are much lower than with the reported realtor fees above. I find several online market places that sell ad space for about €100-€200 with different service levels. They usually provide the customer with required documents, information of the selling process and guidance.

3.1.2 Segmented markets

Real estate markets can be divided into submarkets to allow a more detailed study. As the heading of this section says dwellings, I have already divided the whole real estate market

---

[2] http://www.taloussanomat.fi/tyo-ja-elama/2009/11/21/vedata-valittajaa-saasta-kymppitonni/200924210/139
[3] http://www.hs.fi/talous/HS+selvitti+kiinteist%C3%B6nv%C3%A4litt%C3%A4jien+hinnat/a1305606469642

into submarkets and only chosen housing market. Others include e.g. business premises, public administration buildings and summer cottages. More important division of submarkets for my study is to separate submarkets inside a housing market. For instance, I could take a Finnish city and name it a housing market in a whole, because other cities may have completely different attributes resulting prices and marketing times. To get a better picture of the housing market in this city, I may have to divide it into submarkets if the markets do not seem homogeny. Rest of this chapter only discusses housing markets as it is the focus of my study.

Academics agree that housing markets usually derive of a set of submarkets. The most common way to distinguish a submarket is to divide the market according to geographical areas or characteristics of dwellings (Bourassa et al. 2003). Geographical areas may be chosen e.g. by natural land lines, official city districts or compass directions. Michaels & Smith (1990) find that by following realtors' actions it is possible to find their definition of submarkets in a certain market. They also interview realtors to find their opinion of submarkets and find them to be fairly consistent. Bourassa et al. (2003) argue that hedonic model and multiple regressions are often used for mass appraisal. These methods estimate housing prices pretty accurately and show if a market should considered as a whole or divided into submarkets.

3.1.3 Stable income

What comes to volatility of real estate investments, the income generated are more stable than stock index returns. Figure 7 shows annualized total and unlevered straight residential investment return in Finland from 1998 to 2011. It is divided into income return and capital growth. It is easy to observe, that the income return has remained very constant in comparison to capital growth. In fact, the standard deviation of income return during this period is only 0.4 percent whereas it is 3.2 percent for the capital gain. When compared with stock market and bond returns, one can hold the income return as dividends or coupons on the underlying capital. To understand the low volatility of housing investments, Nyberg (2009) calculates Helsinki Stock Exchange standard deviation to be 32.7 percent and nominal return 18.7 percent from 1912 to 2007.

Total annual return on residential investments is annualized of monthly returns that are calculated with

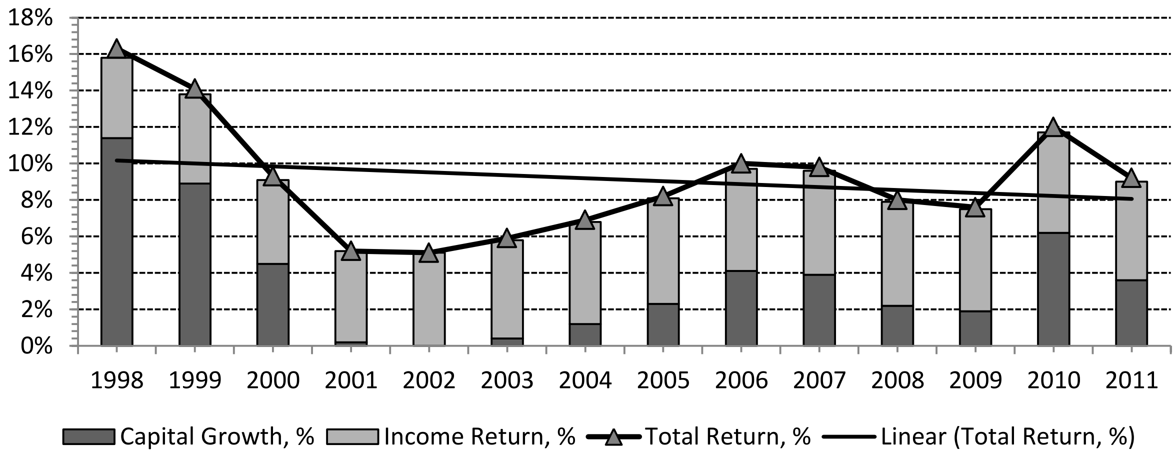
$$TR_t = \frac{\sum(CV_t - CV_{t-1} - Cexp_t + Crec_t + NI_t)}{\sum(CV_{t-1} + Cexp_t)}, \quad (1)$$

where TR_t is the total monthly return of month t and CV_t is the capital value in the end of *t*. $Cexp_t$ are the capital expenditures within month *t* and $Crec_t$ are the capital receipts within month *t*. NI_t is net income within month *t*. To calculate annual return, monthly returns are indexed and the percentage change is observed.

$$AR = \left[\left(\frac{X_{t+12}}{X_t}\right) - 1\right] x100 \quad (2)$$

Figure 7 Annual total return of straight Finnish residential markets investments

This figure shows yields of straight Finnish residential markets as a sum of income return as rents and capital growth as asset appreciation 1998-2011. Annual capital growth and income return may sum imperfectly to total return due to the cross product that occurs when capital and income returns are combined within compounded total return. The data is collected by KTI Kiinteistötieto Oy.



3.1.4 Heterogeneity

Dwellings are heterogeneous and there are no two exactly alike. Buyer preferences determine what type of dwelling she wants and if there is no such dwelling available at the moment, the purchase delays. On the other hand, it may be for the seller that there is no buyer willing to buy her dwelling or her idea of the price differs from the buyer's.

3.2 Pricing

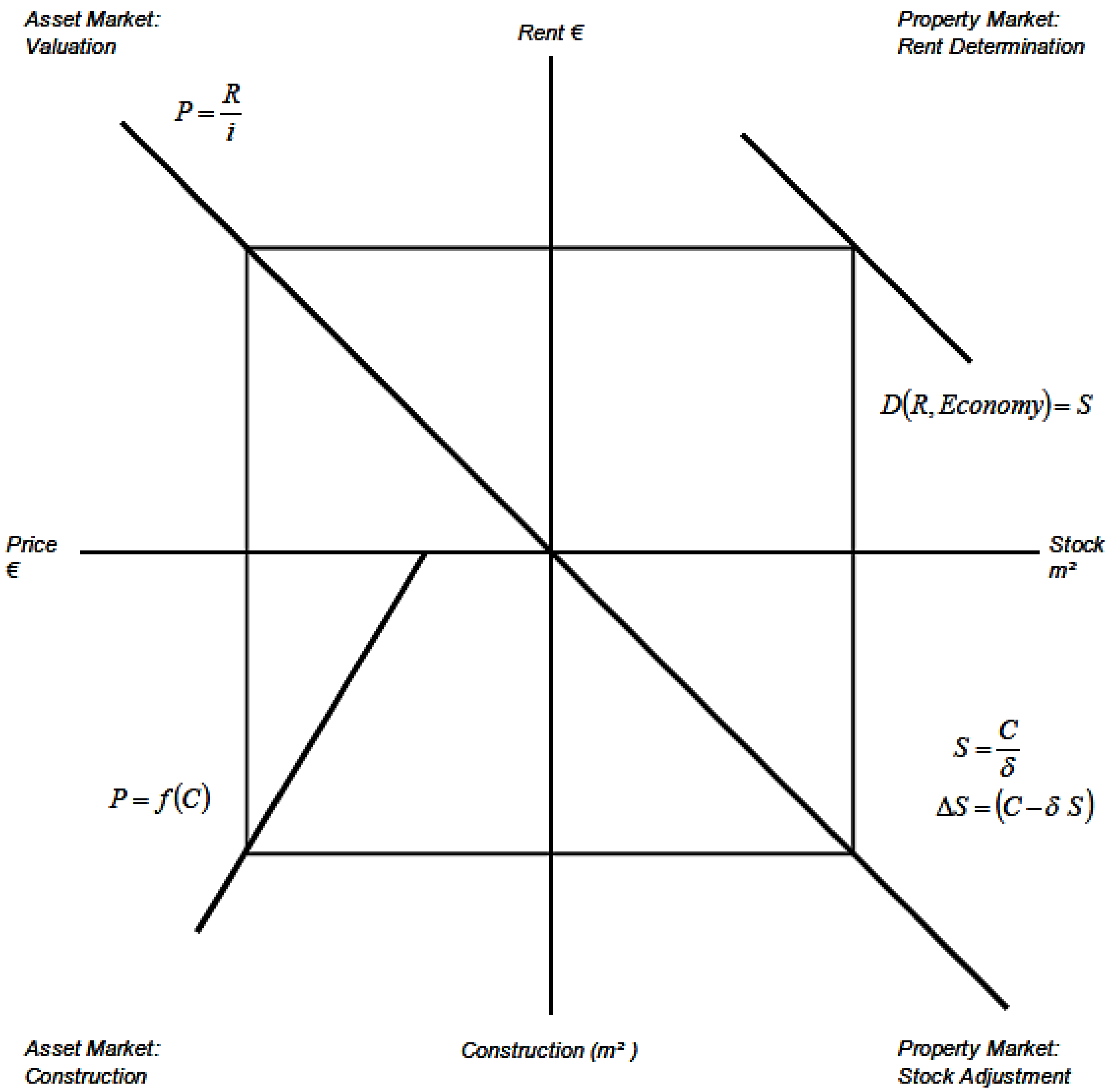
This study is not about pricing of dwellings, but since pricing highly correlates with marketing time I will quickly go through major aspects of it. House price is affected by many macro and micro level factors. Interest rates, employment rate and future forecasts among others affect both the supply and demand. As discussed in next chapter, heterogeneity of houses and people preferences drive the local demand of houses as well as location and buyer's ability to pay. My research question on submarket is about these differences of preferences across a market area.

Hedonic pricing is one widely used tool when considering housing price formation. It explains the value of a house e.g. in one market as a sum of the house's features as estimated from a larger group of houses. These models may predict large portions of the whole value, but leave features hard to put into numbers such as view outside out of analysis. The hedonic model also gives statistical reliability of the results. Sirmans, Macpherson and Zietz (2005) observe the most used independent variables to be floor area, dwelling age, lot size, garage spaces and fireplaces. They base the result on 125 studies on hedonic housing pricing models. As I mention in the introduction, most of housing studies are conducted with US data and this affects the choice of variables in the analyses. Again, heterogeneity and submarkets may lower the actual purchase price if only few potential buyers exist.

Other pricing method include repeat sales model, where price movements are observed from repeated sales of same houses. These models don't provide information on the value of individual house characteristics or on price levels, but have the advantage of being based on actual transactions prices. On the other hand, houses may have gone through renovations which may have changed the value attributes of the house. Malpezzi (2002)

Figure 8 The four quadrant model

The four quadrant model in this figure, introduced by diPasquale & Wheaton (1996), shows the equilibrium in housing market and explains how markets react to shocks both in asset and property market.



Housing markets, like other asset markets, should be in equilibrium in the long run. This means that supply and demand are equal. However, real estate markets are somewhat less liquid than most other markets in the short run. Short term supply is very inelastic due to long construction time of new houses and sticky prices of houses that don't always follow the true value among others. This causes strong short term cycles and enables price development predictability. DiPasquale & Wheaton (1996) explain how the short term equilibrium is found.

In the model the housing markets are divided into asset market and property market. Asset markets derive from asset pricing as shown in the Figure 8. Value of a house here is net rent divided by *i*, which represents interest rate, risk premium and expected increase in rents (as in Gordon growth model, published in 1959). Property markets derive from fact that one area's shortage of housing may not be fully replaced by supply of an area far away. While one area faces incline in prices, may another area suffer from contradictory effect. This would mean there are submarkets within markets that are in my focus as well. This again explains why this model should be used only in a certain (sub) market at a time to see the full effects.

Briefly, the model has following suggestions linking asset market and property markets together:

1) Level of rents affect price level of houses
2) Price level of houses affects the amount of construction
3) Amount of construction affects property stock adjustment
4) Property stock adjustment affects the level of rents

Let's assume that there is a local shock in a market that increases demand on rental housing. Because supply is inelastic in the short term, rents grow and value of properties increases. This signals markets to construct new housing, but since this takes a long time to do and thus to fill the demand, markets remain imbalanced for a while. When new housing is constructed the property stock increases and a new equilibrium is found.

This model also explains the problem of urbanization as mentioned earlier. When demand of rental housing in country side decreases, it lowers both rents and construction. Risk premium increases as the $P = \frac{R}{i}$ slope bends upwards decreasing housing values.

3.3 Studies on time on market

Studies in this field mainly base on applications of micro economical search theory. Search theory studies buyers and sellers who must search for a partner before able to trade. This theory was first used in labor market research and later spread to many fields of research Mortensen (1986). Thus, real estate papers study the search of a buyer and seller for real estate. One key study point here is the time that it takes for those two actors to find each other and to trade. In general, Krainer (1999) says that the time required to sell a house is one of most studied fields in real estate economics and theories have evolved around this question. Most studies on sale time (see e.g. Springer (1996), Asabere et al. (1992) or Kang & Gardner (1989)) investigate the relationship between sale time and dwelling pricing or changes in pricing. They focus in active pricing and how it affects the probability and time to sell.

Older literature on these subjects clearly suffers from lack of data or tools for handling a bigger data set, because their datasets consist of only hundreds of sales. Nevertheless, theoretical models were constructed, but only later they could be tested with a larger sample. Younger studies seem to have rejected many of older suggestion which has to do with larger samples, but also with developed techniques.

Belkin et al. (1976) study 1 000 transactions over a time span of four years to conclude that TOM is an important descriptor of how markets behave, but that house features do badly in predicting TOM. Nevertheless they find some features that are consistently correlated with longer or shorter TOM, which they explain with consistency in pricing errors. They also conclude that differences for various submarkets were so significant that an analysis of submarkets is necessary. In 1974 Cubbin studies a sample of 83 transactions and finds that lower priced houses sold the quicker the higher the price was set. He says it's because buyers reflect the price to quality and don't want to buy a low quality house. Another interesting finding was made by Miller (1978). He presents a theory that with increased TOM, sellers would increase the nominal price of houses to receive a constant real price (inflation was a lot higher than currently). Using a sample of 91 properties he finds promising results, but they were insignificant to support his hypothesis. Kang & Gardner (1989) tested the Miller hypothesis among two other hypotheses and with a larger sample. They cannot confirm the Miller hypothesis, but they find that at times of low interest rates it is better to sell quickly, whereas with high interest rates it is for benefit to wait longer to obtain a higher price. They also find that overpriced and older houses took more time to sell. However, this phenomenon does not occur in case of low priced houses. They interpret that these houses a more of a kind in the first place.

Kluger & Miller (1990) construct a liquidity measure based on Cox proportional hazard technique and analyze their 103 transaction sample with it. This model shows how much more likely is it to get a house sold with any given feature. For example, they show that an additional bed room increases the probability of sale at any given day with 47 percent. They interpret the number to be a measure of liquidity added by the extra bedroom.[4] The writers point out that this model only works with houses sold near market value. Haurin (1988) applies the search theory to study whether more atypically houses have longer TOM than standardized houses. With his sample of 219 transactions he performs a hazard function

[4] The writers have only five independent variables in this model, thus it serves as an example only.

analysis to confirm the research question. Atypical houses may attract fewer buyers and their valuation is more demanding and inaccurate to perform than for a house with lots of other houses to compare with.

Asabere et al. (1992) study the optimal time on market for houses, thus the tradeoff between selling price maximizing and TOM minimizing. They find that both overpricing and underpricing prevent of achieving an optimal TOM. This means that the optimal TOM should come with market pricing. Spinger (1996) studied the effect of seller motivation to TOM. He finds that eager sellers (job loss, divorce etc.) sell at discount, but finds decreased TOM only with foreclosures. This result tells that it is possible to find a house at discount if one can identify the eager seller. Auctions can be used to leverage the price of house sold as high as possible, but costs of setting up the auction may be larger than the gain (Morawski, 2008). Ashenfelter and Genesove (1992) find that in the auction they studied sellers got highest profits for dwellings sold early on in the auction. If the negotiation lasted longer and required face to face discussions the agreed price was 13 percent lower.

Krainer (1999) develops a search model where prices and liquidity are endogenously determined. He defines real estate markets to be hot when liquidity improves with selling volumes and cold when vice versa. One reason preventing hot markets to turn too hot is that sellers like the shorter selling times of hot markets and are willing to sell instead of waiting for a higher bid with a risk that markets would turn cold while waiting. Following the same logic, when markets are cold the sellers do not drop prices to achieve liquidity of hot markets, but wait for the markets to turn. Prices are sticky causing the longer sale times. Genesove and Mayer (2001) explain how the disposition effect on housing markets affects sale times. People facing losses set the price higher than market price thus allowing the sale time to increase. Kaustia (2010) says that there is a strong correlation on housing markets between trading volume and price levels. He argues that the disposition effect might have a significant impact on this correlation. Einiö, Kaustia and Puttonen (2008) study the Finnish markets and the disposition effect as well. With a repeated sales model, they find that number of sales occurring exactly at the original purchase price of the dwelling is unusually high. They also record a significant amount of loss aversion, especially in the greater Helsinki area. The disposition effect helps us to understand how housing markets work and how psychology has an important role on them.

3.4 STUDIES ON SUBMARKETS

Search theory also helps to understand why certain markets may divide into multiple submarkets. Submarket is formed by characteristics of the area and dwellings there as well as by local consumer motivations and behavior. This means consumers in different submarkets may have different attributes they look for and thus never find each other. They may also react differently into economic shocks thus acting differently e.g. with buying decisions.

If two houses in the same submarket are identical or made identical with pricing, they should remain the same time on the market (Belkin et al. 1976). They argue that regressions with house features cannot be used in predicting TOM unless some features of houses are consistently over- or underpriced. Housing markets consist of buyers and sellers (demand and supply) and the more you narrow the market into submarkets, the more this relationship may vary affecting TOM. If a seller promotes a house for sale in a certain submarket, the writers remind that the demand is not all potential buyers in that submarket. Not all are in search of a house at that time and of those who are, only some are looking for the type of house you are selling. The supply consists of other houses in the same submarket appealing to same buyer candidates. Belkin et al. (1976) define submarket to consist of houses that affect each other's prices, thus are competing with each other.

Submarkets are an explanatory factor of many study results (see e.g. Kluger and Norman 1990 or Belkin et al. 1976), but not in the main focus. Pryce and Gibb (2006) made an interesting study where they tried new and more accurate than before methods to analyze submarkets and TOM with Scottish data. New methods account for censoring and duration dependency observations that I will cover in the method section. They show that duration dependency is related to market cycle and submarket structure, but argue that not all implications are known. They suggest all further regression studies to be made with multiple submarkets unlike the earlier ones, as I will do.

Bourassa et al. (2003) find that submarkets have a significant impact on pricing accuracy in their analysis. Not only do submarkets matter, but geography is what makes them matter. "Location, location, location" is not just a tired dictum as they conclude.

3.5 BEHAVIORAL PATTERNS

Third research question of this study investigates behavioral patterns in the data. Finding a clear pattern, let's say that on Monday people list twice as many houses on market than on other days, would raise an interesting question of the causes to this. Behavioral finance is a field of study that usually looks at stock market anomalies on a psychology-based theory approach. It is a study of human fallibility in competitive markets (Shleifer 2000). Aside from the limited arbitrage theory, the other sentiment of behavioral finance is investor sentiment. This is a theory of how investors form their beliefs and valuations or simply how people think like Ritter (2003) puts it. I apply the theory of behaving rational to distribution of housing sell and buy decisions.

Redman et al. 1997 study calendar anomalies with REITs (real estate investment trust) and stocks. These anomalies are January effect, the turn-of-the-month effect, the day-of-the-week effect and the pre-holiday effect. They find positive results for all of these anomalies and motivate me to try these patterns on Finnish markets. I will not measure returns, but amounts of transactions and advertisements. Trading volume creates return at least with stock markets as suggested e.g. by Deo et al. (2008). I don’t find a single paper that would investigate similar patterns as I will with real estate.

I also test the data for search criteria biases. Today it is highly common to search for housing on the internet and online market places allow people to limit the scope of their search. I study whether this search criteria limitation affects the TOM of dwellings. Since consumers limit their searches to match criteria they are looking for, they might lose the options very close to selected criteria thus decreasing probability of sale for those houses. Let’s say a dwelling A is listed with a price of €299 999 and dwelling B €300 000. Dwelling A would fit a search criteria with a price under €300 000, but B not. Dwelling B might be just as appealing to the customer, but is left out of analysis for €1 price difference. If this test succeeds to find statistically significant results this would mean that the seller can affect the expected TOM with cosmetic price setting or have a reason to get the dwelling area re-measured. I try to find published papers of search criteria with other commodities, but unfortunately do not find a single one I could use.

3.6 Event studies

As the name reveals, event studies study the effect of a certain event using adequate data. It is widely used in many fields of research. In finance, researchers measure firm specific and economy wide events such as earnings announcements or change in unemployment rate. (MacKinley 1997) Most of finance related event studies search for price effects around and during an event.

My application of event studies relates to local economic shocks on the housing market. The difference to main stream of event studies is that I don’t look into price effects, but rather focus on TOM of houses and how this is affected by various local shocks such as the local employment rate. In addition, I’m capable of seeing whether shocks affect the number of new listings. A number of academic papers have recorded correlations of economic measures and TOM. In his empirical study Krainer (1999) manages to link TOM of houses with effective interest rate and the job growth rate. He calculates relative odds ratios for 7 percent effective interest rate (1.34) and for 10 percent (0.55). During his study period of 1992-1998 the average effective interest rate is 8 percent. This result means that at the time of 7 percent effective interest rate it is 1.34 times more likely to sell any house than with the average of 8 percent. The average job growth rate was 1.9 percent and given a value of 3.5 percent the relative odds ratio was 1.18.

Anglin (2006) studies how an individual seller reacts to changes in market conditions. As a conclusion he says that all changes in market conditions change either the level or the slope of the price-probability locus; the relationship between the price received and the ease of selling. These changes may force the seller to adjust her selling strategy affecting the outcome. He acknowledges this not to be an empting study, but serves a great supplement to traditional studies that omit market condition measures. A similar type of study was performed by Cheng et al. (2009). They show that sellers with constraints receive a lower price partly due to hurry in the selling process. As they conclude, longer TOM does not simply lead to higher prices. For example Taylor (1999) finds a negative correlation between selling price and TOM. Evans and Lyons (2008) study effects of macro shocks to DM/$ currency ratio over a four-month period. They find that about 20 percent of the total effect variation is due to the news directly and some two thirds of was transmitted via order flow.

4 METHODOLOGY

In this section I will go through methods I use to answer my research questions. I also present the literature using the same methods and rationale behind them. First I introduce the basics of the hazard model and then the applications that I use in this thesis. I end the section with introduction to hedonic regression.

4.1 HAZARD AND SURVIVAL MODELS

Pryce and Gibb (2006) argue that TOM cannot be analyzed as other continuous variables. This is because TOM is a duration-variable and subject to time-dependency and censoring. Time-dependency means that the probability of sale at any given moment may correlate with the time it has already been on the market. The writers argue that traditional approaches such as logit and probit regressions that hold the probability of sale independent perform worse in estimation. They also argue that multiple regressions overlook local dynamics. Other authors have detected the same problem arguing that both of these characteristics (censoring and time-dependency) are overlooked by the widely used ordinary least squares (OLS) and two-stage least squared (2SLS) methods (see e.g. Miller 1978, Kang and Gardner 1989 or Asabere et al. 1993). Pryce and Gibb (2006) say methods that account for these newly recognized characteristics are known as duration models and largely developed in the medical statistics literature. Models can be used for example to compare results of a group given a medicine to another with placebo medicine.

I use hazard models for my first two research questions. Hazard model specifies me the time t probability of an event occurring at time T, given that the event has not yet occurred. Formally, $h(t)$ is the hazard function at time t:

$$h(t) = lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | \mathrm{T} \geq \mathrm{t})}{\Delta t}, \tag{3}$$

where T is the time of failure or retrieval from market in my case.

The survival function shows the fraction of observations that have not yet been retrieved form the market at time T. Kaplan-Meier model allows a straight forward process of estimating the survival function directly from the continuous survival or failure times without noticing the time of the observation.

There are many applications of hazard models used in different papers. These models take housing and other features into account as well as time dependency. Each of the methods specifies a particular shape for the hazard rates. Choosing the wrong model may cause problems with the result and it is better to use more than one model to get reliable results. Krainer (1999) uses three applications that predict the probability of sale depending on the aggregate state of the economy. With these, he studies statistically significant relationships between the marketing times and variables such as employment growth and interest rates. These models may be applied for housing features as well. First there is an exponential hazard that the writer introduces as an easy model to estimate the hazard curve. The second one, the Weibull model, is an application of the exponential hazard that takes time a dwelling has already been on the market into account. Thirdly, the Cox proportional hazard technique compares covariance of two houses with another parameter such as the employment rate as also used by (Kluger and Miller (1990). Next I'll present all these models in more detail and summarize differences of these hazard models with Table 1.

4.1.1 Kaplan-Meier

Kaplan-Meier model allows me to calculate survival rates of marketed dwellings without additional independent variables or time dependency as introduced by Kaplan and Meier in 1958. This method shows the survival rate i.e. the rate of houses not retrieved from market as a function of time and allows censored data.

Let $S(t)$ be the probability that a certain dwelling is not retrieved from market by time t. For a sample of N observations from this population, let the observed marketing times until retrieval of the sample be

$$t1 \leq t2 \leq t3 \leq \cdots \leq tN \tag{4}$$

Corresponding to each t_i is n_i, the number at risk of being retrieved just prior to time t_i, and r_i, the number of retrieved at time t_i.

This model is the nonparametric maximum likelihood estimate of $S(t)$, a product of the form:

$$\hat{S}(t) = \prod_{t_i<t} \frac{n_i - r_i}{n_i} \tag{5}$$

Without censoring, n_i is the number of survivors just prior to time t_i. With censoring, n_i is the number of survivors less the number censored cases. Only surviving cases i.e. those that are at risk to be retrieved from market are observed.

4.1.2 EXPONENTIAL HAZARD MODEL

In the exponential hazard model, the hazard rate is in following form:

$$h(t,X) = \lambda \tag{6}$$

This implies that the model assumes events to occur according to a Poisson function. It means that the risk of failure is flat with respect to time and past events won't have an impact on following failures.

4.1.3 WEIBULL MODEL

The Weibull is a popular generalization of the exponential model, where the hazard rate is characterized as:

$$h(t,X) = \lambda_p(\lambda_t)^{p-1}, \tag{7}$$

where

$$\lambda_i = e^{X_i\beta}. \tag{8}$$

The Weibull model allows for hazard rates vary over time. p is a parameter that determines whether the hazard is increasing ($p>1$), decreasing ($p<1$) or flat ($p=1$) as in the exponential model above. For the housing market purposes I try different values of p, but the optimal model will assumingly have a p-value less than 1 for decreasing probabilities of "sale" over time. I assume this, because houses that sell later often have either a too high price of features potential buyers don't appreciate and thus the probability of a sale with time lowers. Long sale times also drive customers away as they think there is something wrong with the house.

4.1.4 COX MODEL

As Krainer (1999), also Kluger and Miller (1990) use the Cox proportional hazard technique. The hazard function serves a basis to this proportional hazards model as it is also known. Let T be a random variable representing the length of time between the date when the house is put on the market and the date when the house is sold, and let $f(t)$ and $F(t)$ be the probability density function and the cumulative probability density function of sale time. The hazard function, $h(t)$, can be defined as:

$$h(t) \; = \; f(t)/(1 - F(t)). \tag{9}$$

Kluger and Miller (1990) say that the hazard represents the conditional probability of selling a house at time t, provided that the house has not sold until time t. The Cox model is based on the assumption that the hazard functions are proportional and that the proportionality constant depends on the explanatory variables. Thus, if $h(t, X)$ is the hazard function representing the conditional probabilities of sale for a house with explanatory variables X, then the Cox model assumes that

$$h(t, X) \; = \; \exp(\beta X) h_0(t). \tag{10}$$

Beta represents Cox regression coefficients and $h(t)$ is the baseline hazard function. Kluger and Miller (1990) continue that the baseline hazard represents the shape of the hazard function. It can be arbitrarily set to represent the hazard for any dwelling. The hazard function for other dwellings will be proportional to this baseline hazard function.

They note that in this framework, the conditional probabilities of sale for any two houses are proportional, regardless of what time frame we are considering. Thus, if a house in one neighborhood has twice the likelihood of sale in its first week on the market than does a house in another neighborhood, then the first house is also twice as likely to sell in its second week in the market taken that both houses have not sold during the first week. This ratio of conditional probabilities is called the hazard ratio or relative odds ratio. The Cox assumption implies that the hazard ratios do not vary with time allowing us to estimate the proportionality factors without specifying the form of the baseline hazard function. For this reason, the Cox model is often called a semi-parametric method. The Cox model uses a likelihood approach to estimate the vector of beta coefficients from the hazard function. The model first forms the conditional likelihood function by considering the conditional probabilities of sale at each sale time in the sample.

Suppose that at time t_j one home in the sample, with explanatory variables X_j, was sold. Let $R(t_j)$ represent the set of observations with sale or censored times greater or equal to t_j. The probability that the house with X_j sells at t_j, given that we know that one house actually sold at t_j, is

$$h(t, X_j)/\left\{\sum h(t, X_k)\right\}, \tag{11}$$

where k indexes all the houses in $R(t_j)$. Call this probability Π_j. According to the Cox assumption in equation 11, Π_j reduces to:

$$\prod_j = \exp(\beta X_j)/\sum[\exp(\beta X_k]. \tag{12}$$

After this, I can form the conditional likelihood function by multiplying together the conditional probabilities Π_1, Π_2, Π_3 etc. where the index represents each sale time in the sample, with the Π 's adjusted for overlapping sale times. Each Π_j is formed using only the observations that might have been sold at each sale time j. Houses previously sold or censored houses would not be included in $R(t_j)$. The likelihood function is then maximized with respect to the beta coefficients to obtain the Cox model estimates.

Table 1 Hazard models' basic assumptions

This table lists main differences of the three hazard models I use horizontally in the analysis part.

Exponential model	○ Parametric model ○ Events occur following Poisson function ○ Does not take passed TOM into account
Weibull model	○ Parametric model ○ Assumes a monotonic hazard decreasing (or increasing) probability of sale with time passing from market entrance
Cox model	○ Semi-parametric model ○ Hazard function represents the conditional probabilities of sale for a house with given explanatory variables ○ Conditional probabilities of sale for any two houses are proportional

4.1.5 *KERNEL-SMOOTHED HAZARD ESTIMATES*

For the submarket research I follow Pryce and Gibb (2006) and estimate the hazard curve using kernel-smoothened nonparametric estimates of the hazard function for different submarkets in cities of my sample. This model allows me to observe how the probability of sale develops with time. The hazard curves in figures of the results section are based on the following computation of the hazard of sale:

$$\hat{h}(t) = b^{-1} \sum_{j=1}^{D} K\left(\frac{t - t_j}{b}\right) \Delta\hat{H}(t_j), \tag{13}$$

where $\Delta\hat{\hat{H}}$ is the estimated change in the cumulative hazard, *K*() represents the kernel function, $t{-}t_j$ are analysis time intervals, b is the bandwidth[5] and the summation is over the D times at which failure occurs. This nonparametric approach does not constrain the hazard function to follow a particular analytical shape, so I can get a genuine picture of how the hazard function shifts over time and across submarkets.

Pryce and Gibb develop their model by combining graphical approach by Klein and Moeschberger with life tables and likelihood tests of changes to the hazard function. Life tables are derived by grouping data into analysis time intervals given by t_j, where j=1,...,J. Each interval contains the frequency of sale or censoring for the group of dwellings under consideration. The number of dwellings where t_j <=T <t_j+1, where t_j is the analysis time of failure or censoring for property j. The maximum likelihood estimate of the within-interval hazard reported in the life tables is given by

$$h_j = \frac{f_j}{(1-\frac{f_j}{2})(t_{j+1}-t_j)}, \tag{14}$$

where $f_j = d_j/n_j$, d_j is the number of failures during the interval, $n_j = N_j - m_j/2$, N_j is the number of properties still alive at the start of the interval and m_j is the number of censored

---

[5] Bandwidth controls the smoothness or roughness of a density estimate. Pryce and Gibb state that their results are not sensitive to selected bandwidth, but I will test robustness of the model with various values. In general, a low bandwidth does less smoothing of the estimates than a higher one.

observations during the interval. Confidence intervals for the estimated hazards are based on the following standard error suggested by Kalbfleisch and Prentice (2002):

$$s_{hj} = h_j \sqrt{\frac{1 - \{(t_{j+1} - t_j)h_j/2\}^2}{d_j}} \tag{15}$$

4.2. HEDONIC ESTIMATION

Hedonic estimation handles housing as a set of features, a multi-dimensional differentiated good. When estimating TOM, all features have a contribution to TOM of the housing. Following this logic, when selling a house, the expected time-on-market is a sum of housing features, location and other variables affecting this time (Boyle and Kiel 2001). Most papers studying housing pricing use different modifications of hedonic forms to best fit their data and objectives (Laakso 1997). Anglin et al. (2003) also use hedonic regression to better understand their results along with hazard models. They find consistent coefficients with different models, but lower significance levels for the hedonic regression.

Springer (1996) also uses the hedonic regression in his study of single-family housing transactions. He specifies the model for TOM as:

$$lnT_i = B_0 + \sum_{j=1}^{k} B_j H_i + \sum_{j=k+1}^{l} B_j M_i + \sum_{j=l+1}^{m} B_j O_i + e_i\,, \tag{16}$$

where B_i is a vector of model parameters, T_i is the TOM with i as the property, H_i as physical characteristics of property i, M_i represents market conditions at the time the property i was on the market, O_i describes motivations of the seller of property i and e_i is the error term.

5 DATA

In this section I introduce the data I use in the empirical part of this thesis. First I present the housing advertisement data with sources and adjustments to data. Then I continue by introducing other data and descriptive statistics.

5.1 DESCRIPTION OF DATA AND SAMPLE FORMATION

My set of housing data includes observations from three sources; Etuovi, Oikotie and Kiinteistömaailma. In addition, I have data on all houses sold by a certain group of brokers for a limited time span from asuntojen.hintatiedot.fi. Etuovi and Oikotie are both Finnish market leaders as online market places with ca. 45 000 and 35 000 houses on sale at the time of this study. These online market places contain both privately marketed houses and houses marketed via realtors. Yang and Yavas (1994) find no differences in TOM with the choice of a real estate agent. Also the commission rate and size of the agent firm are insignificant in this study. Thus the fact my sample includes multiple real estate agencies should cause no disturbance in the results. Kiinteistömaailma is one of the leading real estate agencies in Finland. Combining the three sources I have a very wide sample of dwelling advertisements in Finland. Observations include dwelling features such as price, size, location and housing form as well as marketing time and the time point of marketing.

Time span of this study is from January 2004 to July 2012 and I have collected altogether ca. 400 000 housing advertisements for the sample creation. I merge the three data sources with creating an identification cell in Excel that includes city, zip code, number of rooms, building type, area, floor and year of construction. I use this identification to erase advertisements of same dwellings present in two or more data sources as duplicates. To control for outliers I erase houses with less than 100€/m2 and prices over €10 million. I also remove advertisements without relevant data such as price, area or marketing time. After I control for overlapping and outliers the main sample consists of 275 304 observations. The data covers six major cities in Finland, namely Helsinki, Tampere, Turku, Oulu, Jyväskylä, Kuopio. These six cities belong to nine largest cities in Finland by population in end of August 2012 (Statistics Finland). The initial data covers dwellings in surrounding towns or cities of Turku, Tampere, Oulu, Jyväskylä that I exclude using the zip codes as determinants. Some of these observations are as far as 70km away from the actual city. For this procedure I use the zip code finder provided by the national postal office Itella[6]. This process is observable in Table 3.

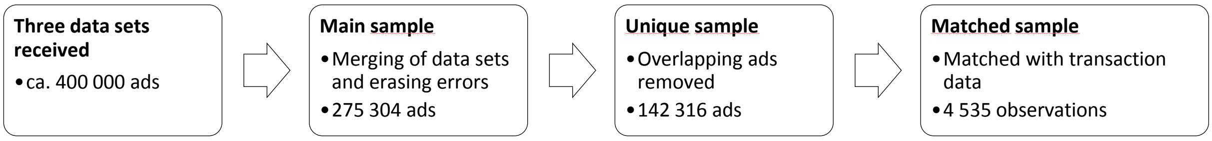
Asuntojen.hintatiedot.fi provides me with housing transaction data allowing me to match advertisements and actual trades. This is the only way I'm able to distinguish transactions from my sample. The service is publicly available and collects data on apartment building and row house transactions from four major real estate agencies. This means I won't be able to match other building types or houses sold by individuals. Thus, the matched sample includes apartment house and row house transactions during the period of July 2011 to July 2012. Time span of this sample is too short for time related variable study, but serves a good basis for market analysis for its period of time. I use cities' official web pages and Google maps to match city districts with zip codes for the Asuntojen.hintatiedot.fi data as it only includes the city district.

I use three different sample variations for this study as presented with Figure 9. First I have all observations after cleaning up the data as the main sample. Then I use a sample that only includes unique observations to control for overlapping advertisements as strictly as possible with the data variables. This means the first sample has similar observations based on the identification. However, with my data, it is possible that I remove unique observations as overlapping advertisements. These are for example similar size apartments in a same apartment house that therefore have the same identification. Finally I use a subsample that I create by matching actual transaction data with advertisements as described above. This way I get a set of transactions with more available data for the observations. However, I don't have censored observations in this sample. This is because I could have transactions my matching process does not recognize during the time span of this sample as well. If I would denote them as censored observation, the model would suffer. Also, if I take all advertisements from this time period, I cannot know if a dwelling is sold by an individual. Secondly, the hedonic model does not recognize censoring and the sample content would change with selection of model. To be able to match all possible transactions in the transaction data, I would need exact addresses or similar unique distinguishers that I do not have and also information on the individually sold housing. These assumptions are in line with Anglin et al. (2003).

[6] http://www.verkkoposti.com/e3/postinumeroluettelo

Figure 9 Sample formation

This figure illustrates the data gathering process and formation of the three samples I use in the analysis. As a notice, all samples include same observations, thus all matched observations are also present in the two other samples.



5.2 DESCRIPTIVE STATISTICS OF HOUSING DATA

In this section I describe the data I use in the empirical part. I start with TOM distribution and continue with observation statistics. Lastly I introduce the process of submarkets creation based on the postal codes.

Figure 10 Frequency distribution of TOM in 2004-2011

This figure shows how TOM of delisted dwellings in the main sample has distributed in weeks from listing from 2004 to 2011. On the vertical axis I count the number of delisted dwellings and on the horizontal axis TOM in weeks. The figure shows data of dwellings that have delisted in one year's time. N=275 304.

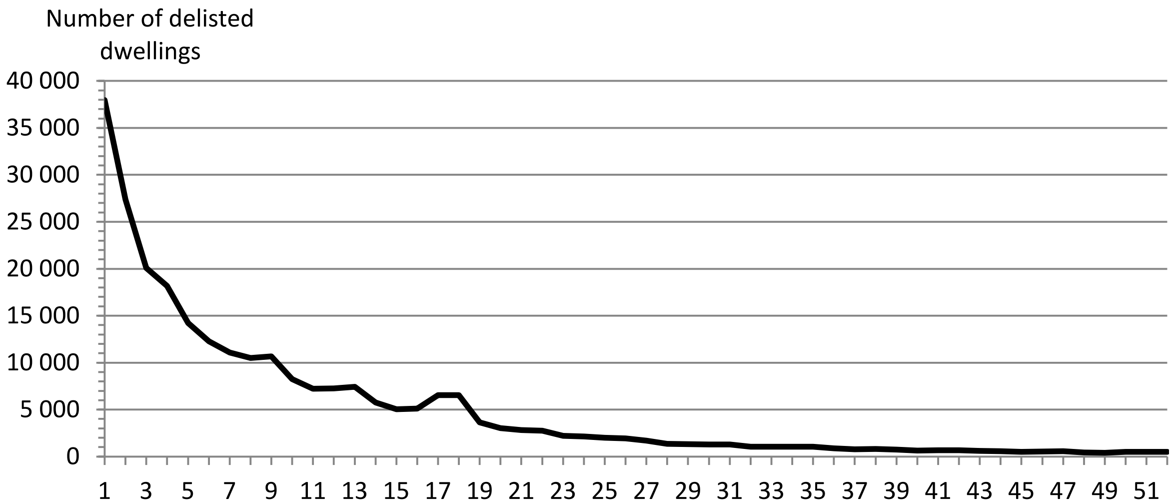
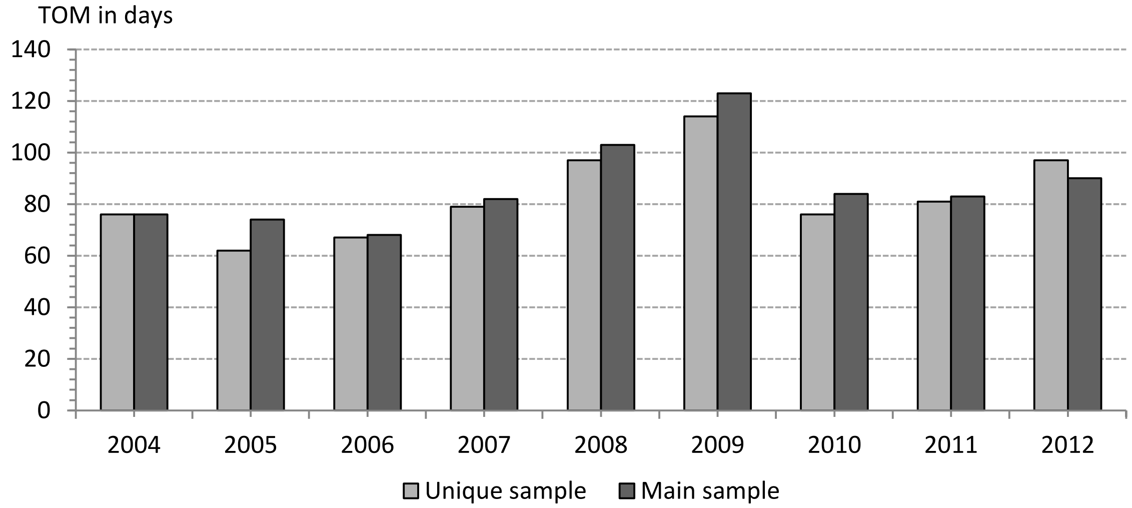


Figure 10 indicates that majority of houses sell relatively fast, within few weeks. As observable, the trend after a year is seemingly flat and therefore the figure data is cut to TOM less or equal to one year. An interesting finding is the high number of housing sold during the first week.

Figure 11 Average TOM in days for sample years

This figure shows the average TOM in days for each sample year 2004-2012, separately for the unique sample and the main sample.



When I compare main sample to unique sample in Figure 11, with about the half of observations, I find the yearly average TOM to be fairly constant between the samples. On average, the unique sample has somewhat lower TOM which could simply be a result of the lower amount of observations, thus lower number of extreme high values of TOM as well.

Table 2 summarizes main descriptive statistics for the main sample and the matched sample observations. This comparison is to identify if the matched sample consists of similar observations than the larger samples or if certain biases exist. As I reason with Figure 11, the larger sample has higher standard deviations and means for all shown variables except for the price per square meter. This variable has increased throughout the main sample time span and since matched observations are from years 2011 and 2012 this result is logical.

Table 2 Descriptive statistics

This table shows means, standard deviations, minimums, medians and maximums for the main sample and the matched sample (in parentheses). Higher variations in values of the main sample are explained both by the sample size and a much longer time period. N=275 304 (4 353).

Variable	Mean	Std. dev.	Minimum	Median	Maximum
TOM	89	127	1	48	2 989
	(57)	(80)	(1)	(28)	(1 322)
Asking price in €	185 916	171 176	10 000	145 000	6 200 000
	(175 322)	(119 163)	(10 468)	(146 900)	(1 750 000)
Area in m2	71	37	10	64	990
	(61)	(25)	(16)	(58)	(269)
Price per m2	2 628	1 350	108	2 338	19 643
	(2 952)	(1 418)	(141)	(2 634)	(10 980)
Year of construction	1975	27	1809	1975	2012
	(1970)	(24)	(1872)	(1970)	(2012)

In Table 3 we see distribution of dummy variables in the main sample compared to matched sample. As I earlier mentioned, Helsinki has a dominant position in both samples. Apartment buildings dominate the building type category which is natural as the samples consist of dwellings in biggest cities of Finland. Both categories go pretty well hand in hand between the two samples. Only distinction is the larger share of Helsinki dwellings, seemingly from Oulu dwellings, in the matched sample that seems to result into higher fraction of apartment houses as well. What comes to other listed dummy variables, the percentage amount of houses with elevators and balconies is on the same level, but the main sample observations tend to have saunas more often in relation.

Table 3 Dummy variables

This table shows distributions of location and season as absolute amounts and relative percentages. Housing type "other" includes wood houses, small apartment houses, separate houses, loft houses and others. N=275 304 (4 353).

	Main sample		Matched sample		
Variable	# of ads	Percentage	# of ads	Percentage	# Zip codes
Location Helsinki	104 081	37.8	1 926	44.2	00100-00990
Location Tampere	48 826	17.7	790	18.1	33100-33900
Location Turku	36 385	13.2	581	13.3	20100-20960
Location Kuopio	19 365	7.0	385	8.8	70100-70840
Location Jyväskylä	22 232	8.2	375	8.6	40100-40820
Location Oulu	44 415	16.1	296	7.0	90100-90800
Building type apartment	203 958	74.1	3 707	85.2	
Building type row house	39 134	14.2	502	11.5	
Building type detached house	15 791	5.8	15	0.3	
Building type semidetached	8 345	3.0	86	2.0	
Building type other	8 076	2.9	43	1.0	
Elevator	112 004	40.7	1 863	42.8	
Sauna	57 334	20.8	322	7.4	
Balcony	149 796	54.4	2 403	49.8	

Table 4 summarizes the time related distribution of the main sample and the matched sample. I define all four seasons to consist of three months as widely used in Finland[7]. Winter includes December, January and February. March, April and May are spring months whereas summer months are June, July and August. This listing leaves September, October and November for the fall period. In Table 5 I list the economic and demographic data I use in the tests to study the housing market characteristics.

[7] http://ilmatieteenlaitos.fi/vuodenajat

Table 4 Distribution of observations in time

This table shows the distribution of sale advertisements in weekdays, months, and years for the main sample and the matched sample.

	Main sample		Matched sample	
Variable	Retrieved ads	New ads	Retrieved ads	New ads
Monday	46 590	23 552	698	439
Tuesday	52 779	42 125	767	573
Wednesday	51 709	53 224	728	744
Thursday	52 409	61 715	821	1 101
Friday	44 651	61 550	654	1 047
Saturday	17 054	26 832	462	370
Sunday	10 112	6 306	223	79
January	23 521	23 341	466	267
February	25 639	23 114	250	121
March	24 343	25 094	157	195
April	22 317	23 605	122	225
May	27 172	29 328	577	749
June	20 763	22 975	355	380
July	19 203	18 851	268	337
August	23 391	23 260	463	466
September	24 767	25 440	433	518
October	23 955	24 407	441	429
November	23 217	22 585	523	409
December	17 016	13 304	298	257
Year -2003		1 179		
Year 2004	11 195	15 051		
Year 2005	26 056	25 966		
Year 2006	32 751	37 246		
Year 2007	40 102	42 362		
Year 2008	40 176	40 879		1
Year 2009	36 092	29 617		3
Year 2010	34 922	36 616		79
Year 2011	37 998	37 172	2 780	3 309
Year 2012 (5months only)	16 012	9 216	1 573	961
Winter	66 176	66 191	1 014	645
Spring	73 832	73 844	856	1 169
Summer	63 357	63 378	1 086	1 183
Fall	71 939	71 969	1 397	1 356
Beginning of month	90 757	96 215	1 510	1 700
Mid-month	89 925	93 061	1 341	1 422
End of month	94 622	86 028	1 502	1 231

5.3 DESCRIPTIVE STATISTICS OF OTHER DATA

I have collected a selected set of economic and demographic data for the six cities under investigation. Table 5 includes these variables with their collection frequencies.

Table 5 Economic and demographic data

This table shows selected economic and demographic data and its collection frequencies. The data comes from Statistics Finland.

Variable	Frequency
Total emigration (net)	Yearly
City to another city migration (net)	Yearly
Within city migration	Yearly
Natural population growth (net)	Yearly
Unemployment rate	Monthly
Average m^2 price in the city	Quarterly

Table 6 shows how the submarkets are formed. I base formation of submarkets on official city practices and categories used by Statistics Finland to get reasonable yet different areas for comparison. City of Helsinki justifies their division principles as follows: Sections should form natural service areas according to residents. Areas should be relatively similar in population and area limits should follow official city district lines. Division should also be permanent to enable comparison in time. I limit my study on submarkets to two major cities Helsinki and Tampere. These two should demonstrate clear answers for my research question and further cities would not bring added value for the study.

Table 6 Description of selected submarkets

In this table I list cities and selected submarkets for my analysis. I divide cities into a reasonable number of submarkets according to official division of districts. Selected submarkets don’t necessarily cover every city district. Zip codes included in selected submarkets are presented in Table 7.

City	Name	Number of advertisements
Helsinki	Southern	24 930
	Western	11 735
	Inner	17 050
	Eastern	16 955
	Northern	17 246
Tampere	Southern	11 371
	Western	8 591
	Eastern	9 555
	Inner	17 000

In Table 7 I show how I put the selected submarkets together. I use the zip code as a selection variable and based compile submarkets based on official city guidelines and own perspective to get different types of submarkets from the two cities. As a notice, I could conduct my submarket study with single zip codes as well, but choose to have larger areas instead.

Table 7 Formation of selected submarkets

In this table I list the zip codes included into selected submarkets. Formation of submarkets follows official city guidelines. Not all zip codes of selected cities are included, as the aim is to produce homogeny samples.

Submarket	Zip codes
Helsinki- Southern	00100, 00120, 00130, 00140, 00150, 00160, 00170, 00180, 00190, 00200, 00210, 00250, 00260
Helsinki- Western	00330, 00340, 00350, 00360, 00370, 00380, 00390, 00400, 00410, 00420, 00430, 00440
Helsinki- Inner	00230, 00240, 00500, 00510, 00520, 00530, 00550, 00580, 00610
Helsinki- Eastern	00900, 00910, 00920, 00930, 00940, 00950, 00960, 00970, 00980, 00990
Helsinki- Northern	00630, 00640, 00650, 00660, 00670, 00680, 00690, 00700, 00710, 00720, 00730, 00740, 00750, 00760, 00770, 00780, 00790
Tampere- Inner	33250, 33240, 33230, 33200, 33210, 33100, 33180,33500,33520,33540
Tampere- Western	33340, 33420, 33330, 33310, 33300, 33270, 33400
Tampere- Eastern	33530, 33560, 33700, 33580, 33730, 33710
Tampere- Southern	33900, 33800, 33820, 33840, 33720, 33850

6 RESULTS

In this section I present the analysis of my data with results. First I test features of the dwellings, their location and other available data to find determinants of time on the market. Then I divide the markets in my sample into submarkets to study the effect of submarket location to TOM. Finally, I will study the data for behavioral patterns. These include calendar and search criteria variables. As this study does not follow given guidelines or underlying papers, I have the possibility to take a closer look on interesting findings and direct the focus on areas of greatest interest. This applies to finding the most suitable form of tests as well. Every test is not highly important because of its results, but they all help finding the most suitable methods and assumptions for studying the Finnish housing market, in the future as well. Based on earlier literature, I cannot form many expectations as they focus on small towns or their submarkets and mainly in the US. These samples also face different regulations and market environments as my Finnish ones. However, certain results are more or less expected and I discuss possible reasons behind them during the analysis.

6.1 DETERMINANTS OF TOM

The first research question is about the determinants of TOM. Based on the variables, samples and selection of research methods I could have countless of different tests to study the markets. I try a lot of different variations and report the ones I feel have the most value for my objectives. As shown later on, even the smallest changes in settings, say controlling the test for a certain variable, can dramatically change the regression results. Next I show the results based on multiple sample and variable set selection as well as different models as presented earlier on. The very basic starting point of my analysis are the Kaplan-Meier estimates that show the fraction of houses not sold after n days from entering the market as plotted in Figure 12.

This simplified figure should give the reader a framework of the approximated time on market of any houses. As we see, I have truncated the time horizon in the figure to 365 days, but already at around 200 days after entering the market, circa 90 percent of all houses are estimated to be sold. In other words, there are houses with long and extremely long selling times, but the majority of houses sell within reasonable times.

Figure 12 Kaplan-Meier estimates

This figure shows the Kaplan-Meier estimates of the matched sample observations (n=4 353) with TOM of 365 days or less. These observations cover 98.87 percent of all observations. This curve estimates the fraction of houses still on the market after n days from entering. This model does not give the exact estimations due to missing censored observations, but works well as a descriptive model.

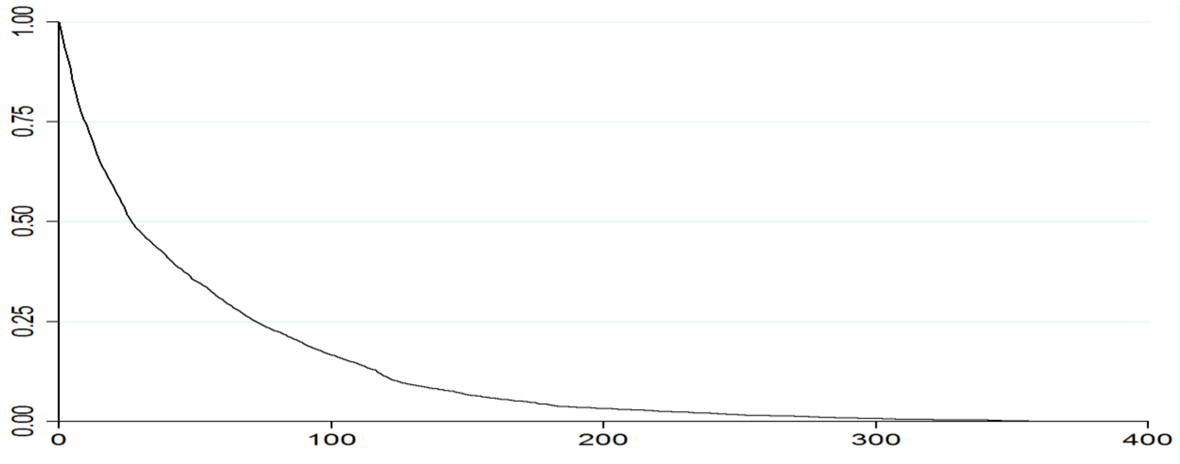


Table 8 demonstrates how the same set of variables run with three different sample settings and two different models affects TOM. This table is to give insight of the dependence of sample selection to results and about the fit of two different models. If I compare the Cox and hedonic models, they seem to give fairly similar results. Only differences in signs come with listing seasons, number of rooms and macro-economic factors. This applies also when comparing different sample results.

Although the literature is strict about the hedonic model's bad fit to housing environment, I show its results, because they are much easier to interpret than those of Cox's model. With hedonic regression, dummy variables mean that TOM is affected by the dummy coefficient value in days, if present with a certain dwelling. With linear variables the interpretation is linear in days of TOM. What comes to hazard models' coefficient interpretation, they measure relative probabilities of sale and the signs are the other way around as for the hedonic regression. This means that a positive coefficient of a hazard model decreases the expected TOM and a negative sign increases it. I run the same regressions with natural logarithm of debt-free price per square meter, but this results lower t/z-values and a lower R^2 values for the hedonic models while having minimal other effects in all cases.

Amount of statistically significant coefficients is about the same with all three samples and the Cox method seems to yield more of them in absolute terms. Larger samples have higher R^2 values as the matched sample. This is due to larger sample size as well as explanatory power of the listing year and macro-economic variables that have limited capabilities in explaining the variation in the matched sample, because of the short time span. What drives the matched sample to have about the same amount of significant coefficients could be that it consists of best observations in terms of sale probability. All observations in this sample are recognized as sold dwellings. Another observation is the low values of the hedonic model's constants. These correlate highly with the independent variable selection as well as data transformation as I explain with Table 9.

For the test shown in Table 8 I have clustered the data according to city of the dwelling. I do this for the macro-economic data that is city related. Without clustering the observations would not function properly as they would compare with other cities' values. My purpose is to study within city changes in these parameters. However, clustering lowers t-values of all coefficients, thus reducing the value of other information and I will not use the macro-economic data in further analyses. I try the test without controlling for the listing year and find the macro-economic data to lose its value in terms of statistical significance in the larger samples while the matched sample remains fairly untouched. This is natural, because the matched sample mainly consists of 2011 and 2012 listed dwellings. With the listing years I witness a doubled R^2 for the main sample and the unique sample.

Table 8 Testing Cox and Hedonic models with different samples

This table presents regression results on determinants of TOM using the Cox model and the hedonic regression methods for the different sample settings. Dependent variable is TOM in days. Figures in parentheses below the coefficients are the t- or z-statistics depending on the model and ***, **, * denote statistical significance at 1%, 5% and 10% levels respectively.

	Main sample		Matched sample		Unique sample	
Variable	Cox	Hedonic	Cox	Hedonic	Cox	Hedonic
Ln (Debt-free asking price)	-0.30***	43.80***	-0.17**	20.15**	-0.36***	45.29***
	(-14.99)	(9.15)	(-2.03)	(3.00)	(-21.14)	(18.40)
Area	-0.01***	0.28***	-0.01**	0.25	-0.00***	0.24***
	(-11.83)	(7.23)	(-2.55)	(0.99)	(-14.73)	(8.50)
(Apartment house)						
Row house	0.19***	-22.55**	0.31**	-34.69*	0.33***	-35.17***
	(2.77)	(-2.63)	(2.45)	(-2.44)	(5.52)	(-5.50)
Detached house	0.42***	-54.80***	0.64**	-53.14*	0.37***	-51.70***
	(5.45)	(-5.63)	(2.21)	(-2.53)	(4.12)	(-4.55)
Semi-detached house	0.19***	-24.06***	0.31***	-22.36**	0.27***	-36.85***
	(3.74)	(-3.15)	(3.32)	(-3.13)	(2.84)	(-5.04)
Loft house	0.21	-2.40	0.28	-19.10	0.08***	-12.00***
	(0.93)	(-1.00)	(1.02)	(-1.29)	(3.72)	(-7.41)
Wooden house	0.94**	-17.62***	-0.39	7.72	0.13***	-22.37***
	(2.13)	(-4.87)	(-1.54)	(0.35)	(4.61)	(-9.60)
(Listing in winter)						
Listing in spring	-0.04*	5.24	-0.04	-0.68	-0.04	5.73
	(-1.74)	(1.75)	(-1.23)	(-0.13)	(-1.15)	(1.31)
Listing in summer	0.05	-3.33	-0.07***	4.42	0.04	-2.81
	(1.20)	(-0.67)	(-2.64)	(1.28)	(0.72)	(-0.53)
Listing in fall	0.08***	-7.87*	0.05	-5.57	0.09***	9.92**
	(2.61)	(2.32)	(1.00)	(-1.14)	(3.98)	(3.69)
(Helsinki)						
Tampere	-0.45***	31.22*	-0.36***	30.42***	-0.20***	22.85**
	(-4.30)	(2.42)	(-9.26)	(8.05)	(-6.40)	(3.15)
Turku	-0.73***	85.90**	-0.29***	20.24**	-0.23***	24.11***
	(-4.72)	(3.40)	(-5.40)	(3.42)	(-4.75)	(4.01)
Oulu	-0.84***	91.21**	-0.27***	34.95***	-0.25***	26.89***
	(-6.38)	(3.29)	(-4.63)	(5.83)	(-10.18)	(10.18)
Jyväskylä	-0.82***	94.21***	-0.32***	34.70***	-0.17***	20.01***
	(-7.83)	(4.92)	(-3.89)	(7.22)	(-17.87)	(11.10)
Kuopio	-0.73***	81.28**	-0.27***	24.42***	-0.13***	13.37**
	(-3.95)	(2.95)	(-3.93)	(6.02)	(-5.07)	(3.15)
(Single room)						
Two rooms	-0.11***	1.41	-0.22***	12.26**	-0.09***	-0.56
	(-2.70)	(0.57)	(-4.48)	(3.02)	(-3.93)	(-0.45)
Three rooms	-0.14***	4.75	-0.22***	14.00**	-0.11***	1.72
	(-3.21)	(1.50)	(-3.37)	(2.85)	(-4.93)	(0.82)
Four rooms	-0.17***	6.32	-0.27***	21.41	-0.09***	-3.29
	(-3.13)	(1.35)	(-2.64)	(2.01)	(-3.26)	(-1.93)
More than four rooms	-0.08***	-2.24	-0.08	14.44	-0.02	-12.31**
	(-2.99)	(-0.60)	(-0.70)	(1.40)	(-1.40)	(-3.48)
Elevator	-0.04	1.54	0.16*	-16.84*	0.03***	-3.94**
	(-0.56)	(0.67)	(1.88)	(-2.28)	(2.92)	(-2.80)
Balcony	0.04	-3.38	0.19***	-15.40**	0.04*	-4.51
	(1.56)	(-0.88)	(3.62)	(-2.72)	(1.77)	(-1.39)
Sauna	-0.11***	4.55	-0.14***	4.43	-0.16***	15.28***
	(-4.32)	(1.76)	(-3.82)	(1.67)	(-14.13)	(11.63)

Table 8 Testing Cox and Hedonic models with different samples (continued)

	Main sample		Matched sample		Unique sample	
Variable	Cox	Hedonic	Cox	Hedonic	Cox	Hedonic
Local unemployment (%)	0.03***	-3.38**	0.01***	-1.16**	0.04*	-4.20***
	(6.07)	(-3.19)	(2.67)	(-2.74)	(7.30)	(-4.18)
Natural population change (%)	-0.23	19.57	0.13	-29.00***	-0.08	2.90
	(-0.54)	(0.48)	(0.89)	(-6.30)	(-0.17)	(0.06)
City to city immigration (%)	-0.92**	123.56*	-0.93*	94.96***	-0.80**	98.52*
	(-2.53)	(2.30)	(-1.92)	(6.43)	(-2.10)	(2.09)
Within city immigration (%)	0.08***	-10.57**	0.00	-0.52	0.08***	-9.53**
	(3.04)	(-3.17)	(0.55)	(-0.96)	(2.98)	(-3.08)
Net immigration (%)	-0.94**	107.60	0.33***	-27.79***	-0.96***	88.11
	(-2.30)	(1.55)	(3.39)	(-4.85)	(-2.75)	(1.69)
Constant	-	-272.37**	-	-220.26**	-	-317.98***
		(-3.71)		(-2.73)		(-7.75)
Year of listing	Yes	Yes	Yes	Yes	Yes	Yes
R^2	-	0.150	-	0.084	-	0.153
N	275 304	275 304	4 353	4 353	142 316	142 316

Next I focus on the coefficients of the test in Table 8. Below I list the independent variables and analyze their meaning and relationship with TOM. I concentrate on the hedonic regression results, as they have a straight forward interpretation. Overall, it is interesting to see that with some independent variable categories one model yields more significant coefficients than the other and the other way around. As I earlier state, a comparison of results to earlier literature is not very informative. However, I compare certain results to papers studying housing prices as the interpretation of some variables is fairly similar: If we say that a certain feature, say sauna, increases the value of a house, it should also increase attraction, therefore shorten the TOM of that house, ceteris paribus.

a) *Ln(debt-free transaction price):* More expensive housing tends to have longer TOM than less expensive ones.
b) *Area:* The linear interpretation is the same as with the price: As a dwelling gets larger, it takes a longer time to sell it. A similar phenomenon is captured by many papers regressing determinants of housing prices, see e.g. Nikola (2011) or Moilanen and Terho (2010)
c) *City:* The city, a dwelling is located in, plays an important part in this test. It is no wonder as the other cities reflect to Helsinki. Buyers and sellers may have different consumption preferences and also the average price levels differ from city to city resulting the coefficients to be significant. In all cases, dwellings in other cities than

Helsinki sell slower. Expected difference can be more than a month when looking from the matched sample results.

d) *Number of rooms:* Here I find differences between the samples, although tendency is that single room housing sells quicker than housing with many rooms. The interpretation itself is also somewhat two-sided. First interpretation is that single room housing sells quicker than a mansion with tens of rooms. On the other hand, large single room housing may have less value to buyers than similar size housing with two rooms. When analyzing the cox results, I find all coefficients for multiple rooms to signal a longer TOM than with a single room dwelling. Most of the coefficients are also significant, so the two models clearly differ with these variables

e) *Building type:* When compared to apartment houses, that represent the majority of observations, all other housing types sell faster. This finding is somewhat surprising, because apartment houses tend to be more homogeny than say detached houses. When testing TOM of different building types for descriptive statistics with the unique sample, I find that apartment houses have the shortest median TOM. However, due to highest standard deviation of results, the average is higher than for other building types resulting to observed coefficients.

f) *Sauna:* All samples indicate that housing with saunas sell slower. As I hold sauna a value feature, it could be that saunas are usually found in more expensive or larger housing and therefore the coefficient is positive for a longer TOM. Nonetheless, the few square meters of space a sauna requires are not necessary to anyone, but something extra in a sense. A buyer can have a feeling she pays too much for these extra square meters if she does not appreciate sauna that much.

g) *Balcony:* A balcony in a house lowers the expected sale time of that house, especially with the matched sample evidence.

h) *Elevator:* Elevator receives contradictory results with different samples. Nikola (2011) or Moilanen and Terho (2010) both find apartments with an elevator to be more expensive, so I think there could be the same phenomenon as with the sauna. In addition, an elevator is more attractive to people living in upper floors, than for those having their apartments lower in the building.

i) *Macro-economic figures:* Focusing in the two larger samples, I expected to find local employment correlating positively with expected sale time, whereas population change and immigration figures negatively. Increase in local employment should lengthen TOM, because of uncertainty leading people to postpone their purchase decisions. Overall, immigration or increase in population should correlate positively with housing market liquidity. However, I find only immigration within a city to follow my logic. Main reason for this is probably the relatively small portions of absolute changes in relation to whole city populations. Thus the coefficients may capture effects I don't see in my data leading to unexpected signs. With employment, another interpretation is that with people losing their jobs need to cut costs and find a cheaper housing to live in.

For the most coefficients I get similar results with all samples and feel comfortable to continue the study with the matched observations. On the other hand, these results allow me to study the other samples in later tests, although transactions are not certain. This gives depth to my analysis as significance levels tend to increase with sample size as I show with Table 8 and time related variation can be dealt with. In Table 9 I study the matched observations more closely, as they consist of additional data I cannot compare with the other samples, such as the transaction price and condition of dwelling. I test the sample with all four methods I introduce in the methods section to see their suitability for this type of analysis. Secondly, I show how dependent the results are on the method selection.

Table 9 TOM with matched observations

This table presents regression results on determinants of TOM using the Cox, Weibull, Exponential and Hedonic regression methods for the matched sample. Dependent variable is TOM in days. Figures in parentheses below the coefficients are the t-statistics and ***, **, * denote statistical significance at 1%, 5% and 10% levels respectively.

	Cox	*Weibull*	*Exponential*	*Hedonic*
Variable	Coefficient	Coefficient	Coefficient	Coefficient
Ln(Debt-free transaction price)	0.26**	0.29**	0.27**	-12.32**
	(2.14)	(2.34)	(2.21)	(-2.05)
Asking price-transaction p. (%)	-0.23*	-0.23*	-0.22*	9.38
	(-1.78)	(-1.76)	(-1.67)	(1.38)
Area	-0.01***	-0.01***	-0.01***	0.49***
	(-5.15)	(-5.32)	(-5.02)	(4.71)
(Condition good)				
Condition satisfactory	0.07*	0.06*	0.06	-3.02
	(1.82)	(1.68)	(1.61)	(-1.59)
Condition bad	0.35***	0.40***	0.37***	-12.57**
	(3.62)	(4.03)	(3.77)	(-2.54)
(Single room)				
2 Rooms	-0.27***	-0.28***	-0.27***	10.98***
	(-4.49)	(-4.69)	(-4.48)	(3.66)
3 Rooms	-0.29***	-0.31***	-0.30***	12.54***
	(-3.19)	(-3.46)	(-3.33)	(2.82)
4 Rooms	-0.29**	-0.32**	-0.30**	12.23**
	(-2.29)	(-2.45)	(-2.34)	(1.97)
4+ Rooms	0.16	0.19	0.17	-5.76
	(0.81)	(0.93)	(0.85)	(-0.58)
(Apartment house)				
Row house	0.06	0.05	0.05	-7.90**
	(0.82)	(0.70)	(0.77)	(-2.07)
Other building type	0.05	0.05	0.05	-5.34
	(0.45)	(0.46)	(0.50)	(-0.99)
Sauna	-0.34***	-0.36***	-0.34***	10.98***
	(-5.05)	(-5.38)	(-5.01)	(3.18)
Balcony	0.04	0.04	0.04	-0.72
	(0.92)	(1.14)	(1.01)	(-0.35)
Elevator	0.06	0.04	0.04	-3.65*
	(1.36)	(0.90)	(0.88)	(-1.74)
Constant	-	-11.00***	-10.27***	1465.61***
		(-6.09)	(-5.71)	(16.31)
Year of listing	Yes	Yes	Yes	Yes
Year of construction	Yes	Yes	Yes	Yes
Zip code	Yes	Yes	Yes	Yes
# Observations	4 353	4 353	4 353	4 353
R^2	-	-	-	0.611
p-value	-	0,93	-	-

If I analyze Table 9 purely as a comparison of methods I conclude the following: All three hazard models yield to basically same results. This goes with both the coefficients and their significance levels. Weibull regression tends to give somewhat higher significance levels as the other hazard models, but this does not apply to all coefficients. With these coefficients I also detect a higher coefficient value. The hedonic model gives same coefficient signs and nearly same significance levels for each variable compared to the hazard models. As I discussed in the literature review, the results were expected to be similar and only small assumption differences between the hazard models result the cosmetic differences. What comes to comparison of this hedonic regression with the one in Table 8, I mainly find similar results, but some differences as well. Introducing additional control variables and other independent variables changes the model outcome.

First coefficient that stands out from Table 9 is the transaction price. With all models, the test indicates a negative relation with price and TOM. This coefficient sign is not affected by selection of asking price or transaction price. I will take a closer look of this phenomenon with Figure 13 later on. Now I focus on the newly introduced independent variables in the model. Discount on selling price, calculated as (asking price-transaction price)/asking price, has a positive correlation with TOM in this test. This result is not expected, because the first thought is that a discount should lead to a quicker sale. However, when thinking about this, other commodities on sale in shopping centers etc. are usually on sale, because they have not sold with initial price.

Applying this theory to housing, a seller should give a discount only if she has waited a long time and wants to get the dwelling sold quicker. But a seller with constraints will set the initial price low to sell quickly. Even so low that many buyers are attracted and the transaction price is actually higher than the listing price, still with a short sale time. Another added variable category is the condition. Compared to housing in good condition, satisfactory and bad condition housing tends to sell quicker. Especially, bad condition coefficient is highly significant with all four methods. Here again, I think this through with support of housing price studies. For example, Nikola (2011) finds that apartments in Helsinki that are in good condition are on average 15 percent more valuable than apartments in bad condition. Therefore, as my results generally suggest, cheaper housing tends to sell quicker. In addition, especially in Helsinki there seems to be a trend of property developing; buying apartments in bad condition to repair them and sell for profit.

In Table 9, I have seemingly high coefficient values. This is because some dummy categories compare correlations with only few observations. If these few observations have a high value TOM on average, the correlations of dummy variables may differ highly from larger samples. As the clearest example I have the earliest year of listing for an observation the year 2008, but as we know the first transaction realized in 24.7.2011. Thus this observation must have a very long TOM and since all other observation reflect to this, the coefficients of year of listing for other years with more observations are large as well as the constant. Simply by reducing year of listing in Table 9 I reduce the constant of the hedonic regression from 1465.61 to 379.30 while maintaining all other coefficient signs and significance levels. As many as 76 percent of matched sample observation advertisements are put on the market in year 2011. In further tests I aim to control for these unpleasant coefficients with grouping of dummies as later explained. However, as my dependent variables cannot all be zero at the same time with it, the constant does not have an intrinsic meaning.

In Table 10 I test the unique sample with different specifications of variables. I use the hedonic regression for this analysis for easier interpretation and comparison of scenarios. In addition to single coefficients, I'm interested to compare R^2-values of the different scenarios. As I earlier explain, the selection and formation of dummy categories is essential to test results. With tests in Table 10 I have grouped the year of constructions into decades to reduce variation in the model caused by single years. In the second specification I transform the dependent variable into natural logarithm. For the independent variables I have two specifications as well: (1) all linear or dummy variables and (2) all natural logarithms or dummy variables. These specifications lead to four scenarios presented in Table 10.

The results are generally highly significant and yield fairly similar R^2 values. The only differences of coefficient signs come with sauna and the constants. The differentiating sauna coefficient is not significant, but an interesting finding is that the constant signs are positive for the specification (1), but negative for specification (2). This seems to derive from lower price and area values' range or standard deviation with specification (2) as they are in natural logarithm form. This results to higher coefficients and therefore to lower constants as each euro and square meter of area increases the TOM. What comes to other dummy variables, they are constantly lower for specification (2) than for specification (1). The interpretation derives from the linear variable reasoning; as the range of linear variables increase, the dummy values increase to capture similar size effect than for lower linear dummy variables with specification (1). Specification (2) yields higher R^2 values which I interpret to result from variable ranges to be more comparable. By this I mean that if I have, say a range of 1-200 for TOM and 10 000-2 000 000 for price, these categories are less comparable than if the price range is transformed into natural logarithm.

Table 10 Determinants of TOM with variable transformation

This table presents regression results on determinants of TOM using the hedonic regression for the unique sample. Independent variables for the two specifications are: (1) all linear or dummy variables, (2) all natural logarithms of variables or dummy variables. Figures in parentheses below the coefficients are the t-statistics and ***, **, * denote statistical significance at 1%, 5% and 10% levels respectively.

	(1)	(2)	(1)	(2)
	TOM as dependent variable		ln(TOM) as dependent variable	
Variable	Coefficient	Coefficient	Coefficient	Coefficient
Debt-free price	0.00***		0.00***	
	(25.07)		(15.10)	
Ln(debt-free price)		34.32***		0.23***
		(39.78)		(22.52)
Area	0.25***		0.00***	
	(17.62)		(22.74)	
Ln(area)		35.15***		0.63***
		(22.21)		(34.12)
Number of rooms	3.13***		0.09***	
	(8.15)		(19.73)	
Ln(number of rooms)		-19.10***		-0.14***
		(-14.31)		(-8.91)
(Apartment house)				
Row house	-23.99***	-30.99***	-0.26***	-0.32***
	(-20.55)	(-25.55)	(-18.87)	(-23.26)
Detached house	-27.84***	-38.24***	-0.22	-0.31***
	(-17.93)	(-25.42)	(-12.38)	(-17.79)
Semi-detached house	-27.84***	-36.67***	-0.23***	-0.30***
	(-15.26)	(-20.11)	(-10.70)	(-14.49)
Other building type	-12.56***	-11.84***	-0.06***	-0.05**
	(-7.01)	(-6.63)	(-2.84)	(-2.52)
Sauna	1.73**	-0.37	0.88***	0.07***
	(2.29)	(-0.05)	(9.85)	(7.43)
Balcony	-10.25***	-12.04***	-0.05***	-0.08***
	(-15.02)	(-17.64)	(-6.39)	(-9.74)
Elevator	-5.74***	-5.75***	-0.07***	-0.08***
	(-7.75)	(-7.80)	(-8.51)	(-9.06)
Constant	132.39***	-378.00***	3.71***	-1.02***
	(44.28)	(-39.05)	(104.81)	(-8.94)
Year of listing	Yes	Yes	Yes	Yes
Decade of construction	Yes	Yes	Yes	Yes
Zip code	Yes	Yes	Yes	Yes
Number of Observations	142 316	142 316	142 316	142 316
R^2	0.199	0.210	0.182	0.193

However, studying all cities as a whole may give a wrong picture of the Finnish housing market. This is due to large proportion of apartment houses and houses in Helsinki as well as price differences between various cities. Also, by grouping observations to otherwise more homogeny samples may give valuable information that is not recordable from a diverse sample. I continue my study with Table 11 by dividing the sample into cities and asking price ranges to see whether the results get more consistent. This procedure is also supported by Springer (1996), who divides his sample to different price categories for consistency.

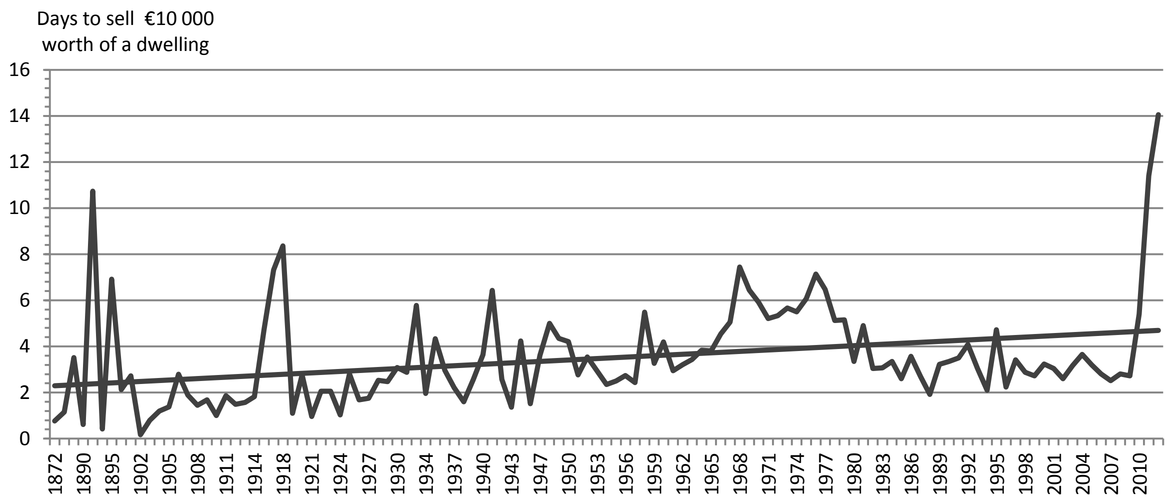
The first conclusion from Table 11 is that results vary a lot and have very few statistically significant coefficients. This finding is not in line with my assumptions of less variation with more focused samples. One possible reason for this, again, is the reduced sample size. On the other hand, especially Tampere and €100k-€200k category show higher model fit than in most other tests. This indicates that focused samples enable to explain the local variation better than with less focused samples. One interesting finding is that the coefficient for price discount follows my initial logic with all samples other than Helsinki and is also significant for Kuopio, Tampere and €100k-€200k price category. Area coefficients are also dependent on the sample selection, although only positive coefficients are significant. Another interesting finding, although mainly insignificant, is that additional room coefficients are all negative in the €100k-€200k price range, whereas all are positive with €200k-€400k sample. An interpretation of this could be that if a lower priced housing sells cheap, it is either a cheap housing in general or the area is very efficiently in use attracting buyers willing to pay for needed space only. Other coefficients have same signs, except of listing season that has contradictory, but insignificant results.

I find that when controlling for the year of construction, the coefficient for debt-free housing price is negative for all samples except the €200k-€400k and insignificant for all except Kuopio. When I test the same samples overlooking the year of construction, I get a positive coefficient for all except Kuopio and all of them are significant at least at 5 percent level. To capture the effect of construction year to price coefficient I test the matched sample controlling for selected ranges of construction year. As I divide the sample according to construction year; -1929, 1930-1959, 1960-1989 and 1990-2012 I find only the newest sample to have the positive coefficient for the dwelling price.

To fully understand this relationship I calculate average time to sell €10k of any housing in the matched sample and plot them to Figure 13. This figure indicates that the relationship of TOM adjusted with dwelling price to construction year varies a lot according to the latter. Therefore I conclude that the sign of the dwelling price coefficient is fully dependent on the sample selection. However, as the trend line indicates, for the whole sample the relationship is rising Thus I conclude the newer the building the longer it takes to sell a dwelling of the same price. The peak in the figure with years 2011 and 2012 has two reasons I can think of. Firstly, Figure 5 suggests that there was a slight peak in new construction from 2009 to 2012. Similarly, marketing times remained constant as seen from Figure 11 and excess supply occurred. Also, new housing is generally more expensive than older housing and with speculation of the economy, this probably lead people to make less risky transactions with lower debt requests and such. Especially in Helsinki, many of the newer building are built in expensive areas. This feature could support the peak in Figure 13 also. Rakennuslehti (11.10.2012), a Finnish construction magazine, states that building of extreme high priced buildings has declined and the supply returned to normal level[8]. Interesting information is that the share of new build housing of all transaction is currently at highest in ten years. However, I also test the same figure excluding years 2010-2012 and the trend line remains clearly upwards.

Figure 13 Relationship of TOM, dwelling price and construction year

In this figure I plot the price adjusted TOM for each construction year. The value on the Y-axis represents how many days it takes to sell €10k worth of any housing. For example a dwelling costing €100k that was built in 1988, took 20 days to sell on average during the matched sample period. Straight line in the graph plots the linear trend.



[8] http://www.rakennuslehti.fi/uutiset/lehtiarkisto/29615.html

Table 11 Determinants of TOM with selected city markets

This table presents regression results on determinants of TOM using the Hedonic regression for the subsamples formed from the matched sample. Figures in parentheses below the coefficients are the t-statistics and ***, **, * denote statistical significance at 1%, 5% and 10% levels respectively.

	Helsinki	Kuopio	Tampere	€100k-€200k	€200k-€400k
Variable	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Ln(debt-free transaction price)	2.88	-68.78***	-45.48**	-25.54*	14.20
	(0.25)	(-2.77)	(-2.47)	-1.92	(0.48)
Asking price-transaction p. (%)	0.36***	-0.45*	-0.37*	-0.11	-0.04
	(2.72)	(-1.76)	(-1.87)	(-1.00)	(-0.42)
Area	-0.06	2.56***	0.90**	1.36***	0.11
	(-0.33)	(3.73)	(2.40)	(5.59)	(0.48)
(Condition good)					
Condition satisfactory	-1.84	-3.93	2.53	-0.06	-7.25
	(-0.47)	(-0.44)	(0.39)	(-0.05)	(-0.65)
Condition bad	-14.90**	-23.89	-35.13	-9.70	-18.90
	(-1.96)	(-0.67)	(-1.58)	(-1.12)	(-1.03)
(Single room)					
2 Rooms	15.76***	10.34	17.54*	-2.38	19.71
	(2.92)	(0.68)	(1.94)	(-0.46)	(1.16)
3 Rooms	23.46***	5.62	16.70	-17.53	28.68
	(3.00)	(0.25)	(1.18)	(-1.43)	(1.42)
4 Rooms	37.78***	-17.02	40.89*	-17.34	48.86**
	(3.32)	(-0.60)	(1.95)	(1.43)	(2.03)
4+ Rooms	30.60*	-62.87	-7.77	-31.80	42.30
	(1.77)	(-1.36)	(-0.21)	(-1.28)	(1.37)
(Apartment house)					
Row house	-15.76*	-22.89*	-7.84	-11.93*	-25.69*
	(-1.91)	(-1.81)	(-0.76)	(-1.78)	(-1.78)
Other building type	-22.54**	-1.21	-10.50	-13.20	-20.03
	(-2.21)	(-0.04)	(-0.71)	(-1.23)	(-1.29)
Sauna	14.55*	12.16	7.65	4.35	10.03
	(1.91)	(1.22)	(0.85)	(0.91)	(0.97)
Balcony	-7.08*	-23.18***	-0.45	-6.88**	-5.75
	(-1.85)	(-2.97)	(-0.07)	(-2.04)	(-0.64)
Elevator	-8.55**	-5.91	-7.19	-2.70	-27.16***
	(-2.26)	(-0.66)	(-1.20)	(0.66)	(-3.14)
(Listing in winter)					
Listing in spring	4.41	5.92	-2.18	-5.52	3.90
	(0.86)	(0.54)	(-0.28)	(-1.29)	(0.37)
Listing in summer	4.48	-1.04	4.60	4.66	-1.32
	(0.88)	(-0.10)	(0.59)	(1.08)	(-0.13)
Listing in fall	-2.42	7.93	-17.19**	-4.04	-11.38
	(-0.50)	(0.73)	(-2.20)	(-0.97)	(-1.13)
Constant	139.64	775.49***	651.11***	385.26**	26.60
	(1.02)	(2.81)	(2.97)	(2.43)	(0.07)
Decade of construction	Yes	Yes	Yes	Yes	Yes
Zip code	Yes	Yes	Yes	Yes	Yes
Number of Observations	1 926	385	790	2 265	932
R^2	0.225	0.254	0.539	0.402	0.286

6.2 SUBMARKETS

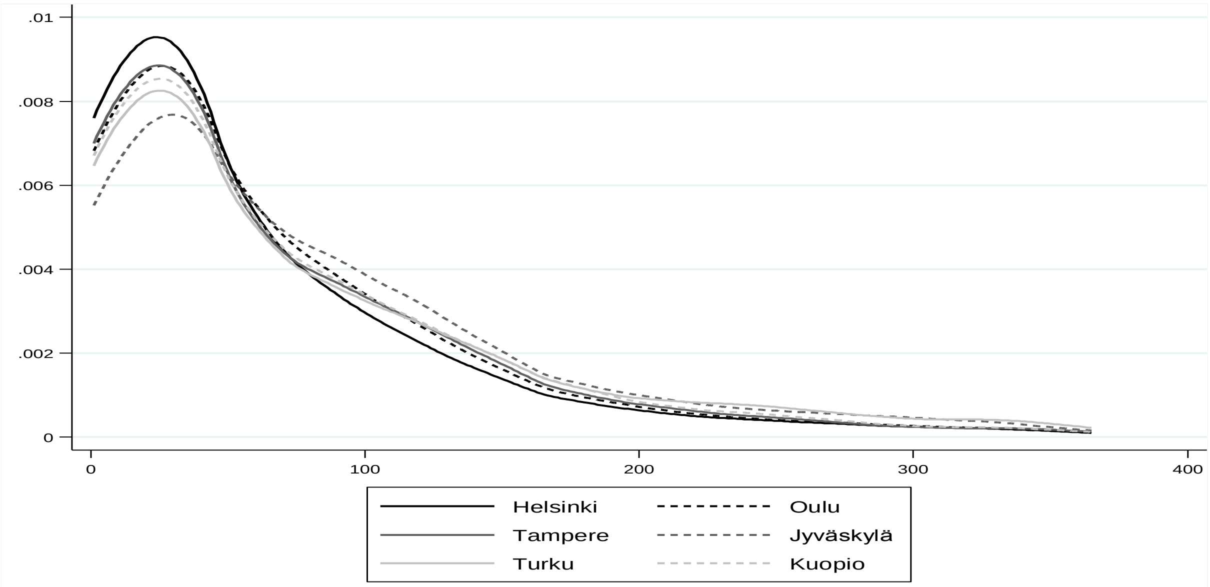
A logical way to continue the analysis of the Finnish housing market is to study split city markets into smaller submarkets. Not only, because above I show that there is need for closer scope in analysis, but also to see if there is variation even within cities. I also study submarkets during different economic states to see whether they have impact on the TOM estimates and if different submarkets react differently to economic changes. I use kernel-smoothened hazard estimates that are basically histograms of hazard estimates smoothened for easier comparison. The higher the hazard rate, the higher the probability of sale at n days after entering the market.

As this analysis is descriptive in nature, I use the unique data set. A significant reasoning to choose unique observations is also the larger amount of estimates for each submarket. For the following figures I have capped the days of TOM into one year. This is because a longer tail has no valuable information and this way the first year of TOM is better visible. In Figure 14 I plot the unique sample for the whole study period by city market to see variation between them. As we observe, dwellings in Helsinki have higher probability for sale during the first weeks than others. This finding is in line with the results of previous regression tables. Lowest hazard rates in the beginning are found with Jyväskylä dwellings. As this is a density function, I observe a higher probability of sale after 100 days in the market for Jyväskylä dwellings whereas Helsinki is at the bottom to offset the initial difference. This is due to amount of sold houses in Helsinki before this time point in relation to Jyväskylä.

In Figure 15 and Figure 16 we see how these submarkets I form behave during three different years. There are two interesting things I study with these figures. Firstly, I compare submarket curves to each other: which ones seem to suggest a shorter TOM than the others. In addition I'm interested to see whether submarket liquidity is constant over time or if some submarkets react to environmental changes more heavily than others. Secondly, I look at the shape of the curves, the highest hazard rate the estimation yields and other hazard rate values for example at 100 days. The higher the curves peak, the steeper they decline after the initial peak. This suggests a shorter average TOM for that period. Following the same logic, if the initial peak is mild, the right-hand side tail is fatter and the average TOM increases.

Figure 14 Kernel-smoothed hazard estimates for whole city markets

This figure plots the hazard estimates of retrieved advertisements for each day a dwelling is on the market for the whole sample period 2004-2012. Here I have combined submarkets of each city to see if different city markets act differently. Bandwidth is set at 20.



With a closer analysis of Helsinki and Tampere submarkets I find interesting results. In Helsinki, I would have expected more variation with different submarkets, but still the curves contain interesting information. From the shape of the curves I see that in 2008 the average TOM for each submarket was longer than in the other years. This was a time of economic downturn, which probably caused this effect. In 2005, northern Helsinki dwellings had the shortest average TOM and western Helsinki performed the slowest, based on the curves. In 2008, inner Helsinki dwellings sold the quickest with eastern Helsinki being the slowest. By 2011 the situation returned to same as for 2005. An interesting finding is that the curves for western Helsinki have remained fairly constant over time maybe indicating less variation with time or economic state in general on this market. Inner Helsinki dwellings seem to work similarly, but eastern Helsinki suffered in 2008 situation to return to 2005 level in 2011.

Tampere submarkets seem to function more differently from each other than in Helsinki. In 2005, the western submarket is the only one standing out from the others with longer selling times. Other than that, submarkets work fairly similar in relation to each other’s over time. With Tampere in 2008 I get the most differences between the submarkets in this analysis. Compared to Helsinki, I conclude this is due to 1) a higher correlation with the economic state or 2) the smaller sample size.

To sum up, this test is not highly informative with these submarkets compared to city markets. Perhaps the city markets are that integrated that no clear boundaries exist. With a closer division of areas, say to zip code level, the results could have higher variation. Simply, my results suggest the TOM is fairly little affected by my division of submarkets compared to corresponding city markets. For the curve shifts I can think of multiple reasons. By comparing the curves to macro-economic data I find, for example, that the unemployment rate follows nicely the shape of the curves. In 2005 the unemployment rate was above 9 percent, whereas in 2008 at about 6 percent and 7 percent in 2011.Looking at the net immigration and natural population increase I find similar results. For Helsinki, the net immigration is nearly 10 times higher for years 2008 and 2011 than for 2005.

Figure 15 Kernel-smoothed hazard estimates for Helsinki submarkets

In this figure I plot the hazard estimates of Helsinki submarkets as presented earlier. I use years 2005, 2008 and 2011 to see possible changes in relation with different time points. Bandwidth is set at 20.

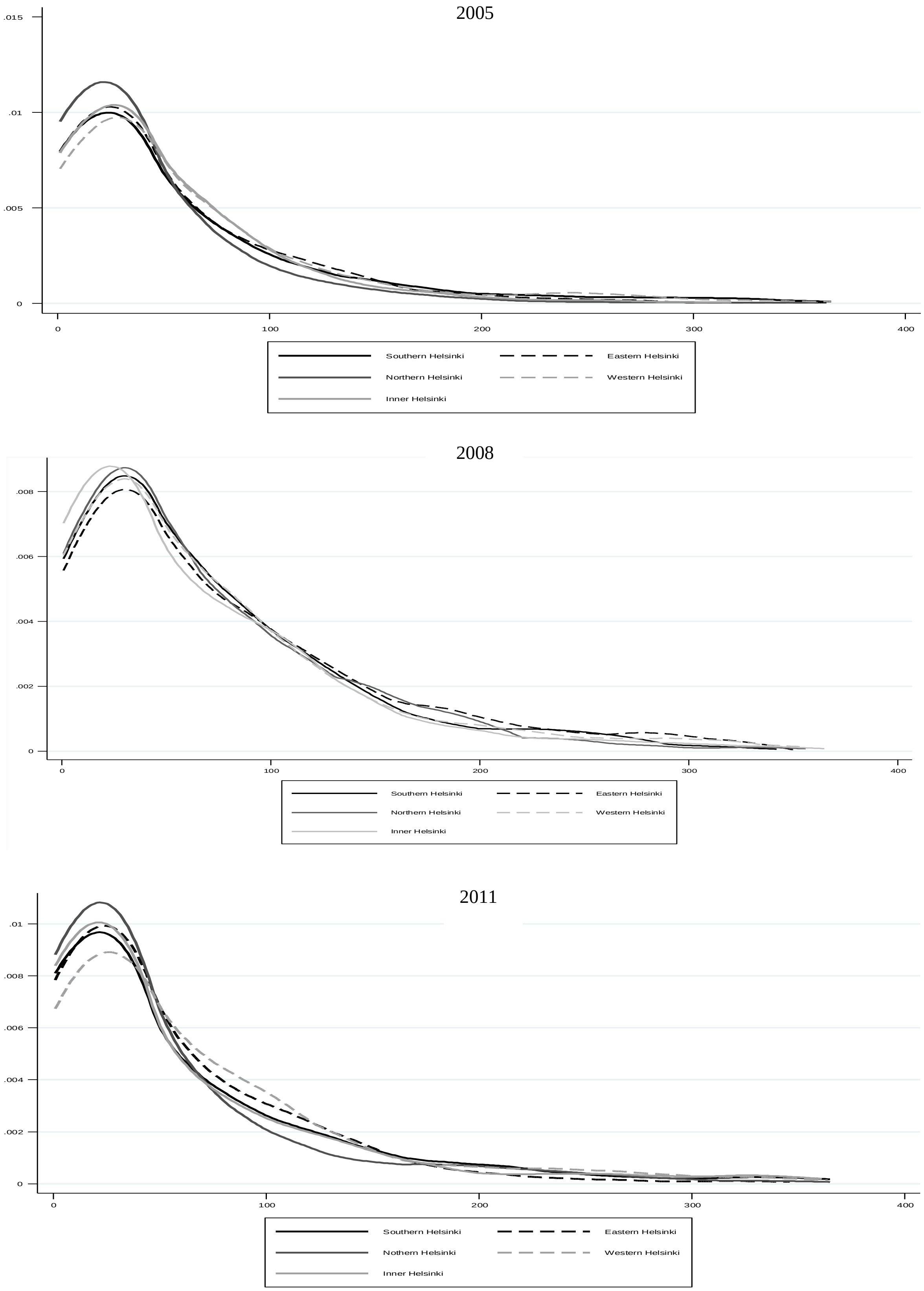
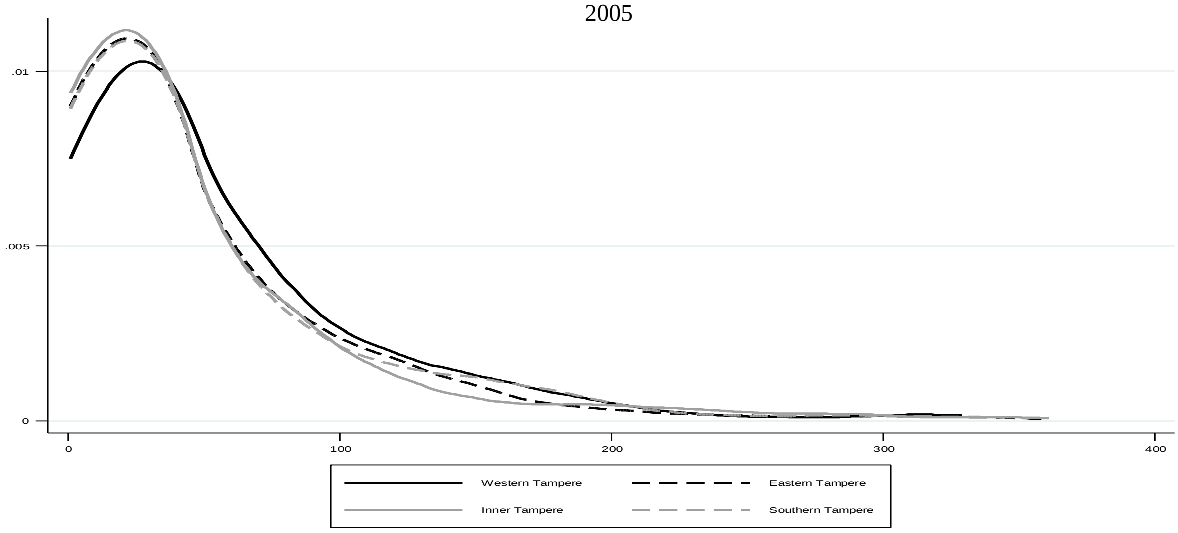
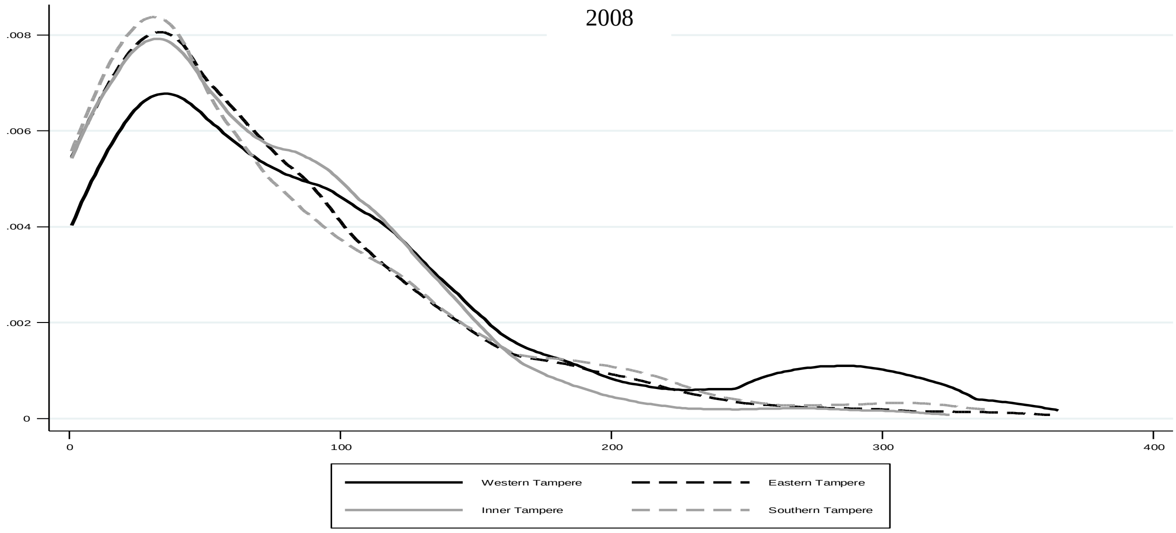
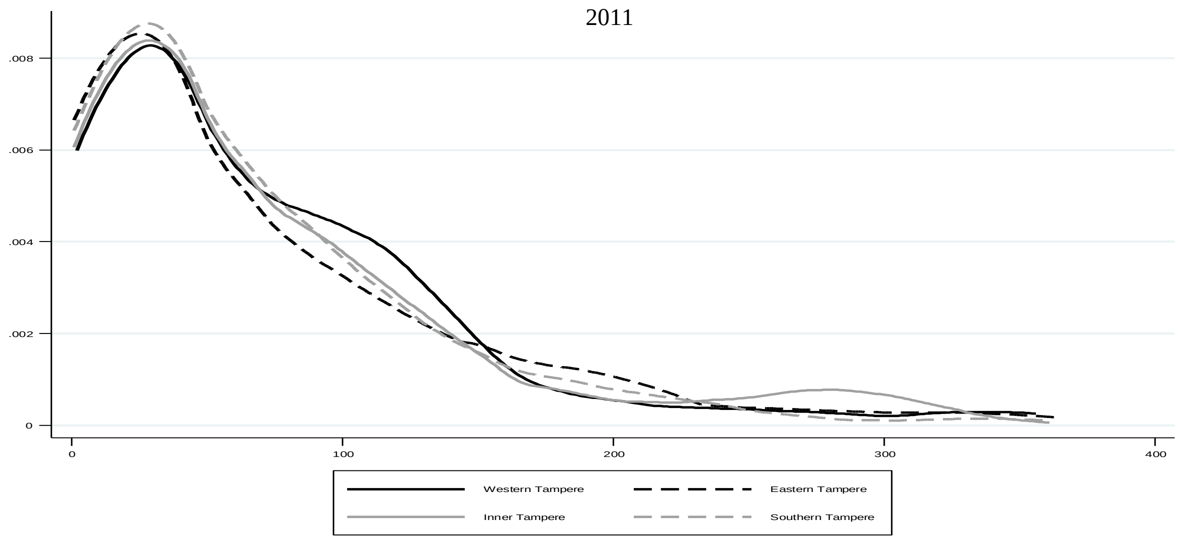


Figure 16 Kernel-smoothed hazard estimates for Tampere submarkets

In this figure I plot the hazard estimates of Helsinki submarkets as presented earlier. I use years 2005, 2008 and 2011 to see possible changes in relation with different time points. Bandwidth is set at 20.







6.3 Behavior

In this section I present the results for behavior related studies on TOM. First I go through calendar related questions followed by a study of search criteria effect on expected sale time.

6.3.1 Calendar

With other assets, academics have recorded various behavior related patterns, usually concerning asset returns. I test selected anomalies with advertisement entry dates representing beginning of sale time, to find differences in expected sale time rather than asset return. The purpose is, again, to describe the Finnish housing markets as any results should be fresh and interesting to readers. I test the calendar effects with all three samples. As Table 12 suggests, most calendar related coefficients are statistically significant. The matched sample coefficients are biased to years 2011 and 2012, whereas the two larger samples have a time span of at least ten years. Therefore, contradictory signs between the matched sample and others are not crucial now, since the matched sample describes the situation for a different time span. Main sample and the unique sample have three contradictory signs, from which only December if statistically significant for both. Thus the results should describe the Finnish market behavior fairly well for the last decade.

Starting with the weekdays, all samples name Thursday to be the best day to list a dwelling in terms of expected sale time, although this is not statistically significant. Monday has been the second best day to enter the market with the main sample and the matched sample, whereas the unique sample suggests this to be Sunday. The worst performing day in this analysis is seemingly Tuesday. The matched sample results suggest that during 2011-2012 the worst entrance day would have been Sunday. Despite the small differences between the samples, the two larger ones suggest that differences in expected sale time are more than a week with a careful selection of the market entry day. When it comes to the most optimal month for listing a dwelling, it has been the late summer months or April and May during the last two years. Also the year change seems to be a rather good time point for listing. As most of the coefficients are significant, I conclude that a seller with no pressure to sell should wait for an optimal time to list her dwelling as the results suggest differences in expected sale time to be weeks, even a month between various listing months. The optimal time indicates a "hot market" suggesting higher interest towards any housing ceteris paribus. However, the situation may change on a yearly basis according to other factors than calendar relates as well and one must not blindly follow these types of suggestions. Breaking months into three parts I find only a couple days differences, but overall, middle of a month has been a better time to list than earlier or later in a month. I use the listing year for the larger samples as a control variable only, as the average TOM for different years is publicly available information.

It is rather hard to reason why a certain day is better for listing than another, especially with a commodity that on average takes many weeks to sell. It could be that certain type of sellers list on different days, with more or less selling pressure or constraints. With months, I expect certain seasons to be more liquid than others. This variation in season liquidity could correlate with other asset markets. Perhaps investors liquidate their other assets during the peaks of housing market liquidity, credit is easier available or other similar reasons lead to this result. With all samples, R^2 remains at around 3 percent level indicating a poor forecasting power with these variables alone.

Table 12 Behavior of TOM according to advertisement put date

In this table I show results on calendar behavior. All independent variables in the model are dummy variables, thus I have selected Monday, January and beginning of month as benchmark variables. Figures in parentheses below the coefficients are the t-statistics and ***, **, * denote statistical significance at 1%, 5% and 10% levels respectively.

	Matched sample	Unique sample	Main sample
Variable	Coefficient	Coefficient	Coefficient
(Monday)			
Tuesday	5.86	7.93***	8.82***
	(1.16)	(5.91)	(8.67)
Wednesday	0.51	3.16**	3.24***
	(0.11)	(2.46)	(3.31)
Thursday	-5.20	-1.73	-1.47
	(-1.16)	(-1.38)	(-1.54)
Friday	10.84**	3.60***	3.67***
	(2.39)	(2.88)	(3.82)
Saturday	2.33	4.01***	6.19***
	(0.43)	(2.74)	(5.54)
Sunday	24.03**	-1.57	3.46*
	(2.48)	(-0.68)	(1.96)
(January)			
February	29.77***	7.70***	8.46***
	(3.42)	(4.96)	(7.28)
March	30.94***	5.73***	4.89***
	(4.13)	(3.80)	(4.30)
April	-20.03***	3.71**	2.72**
	(-4.82)	(2.43)	(2.36)
May	-13.42**	9.38***	6.45***
	(-2.35)	(6.48)	(5.88)
June	14.96**	17.99***	15.11***
	(2.36)	(11.70)	(12.88)
July	2.57	-6.82***	-5.60***
	(0.40)	(-4.19)	(-4.62)
August	-3.06	-9.57***	-15.85***
	(-0.50)	(-6.27)	(-13.59)
September	-0.84***	0.68	-2.24**
	(-0.14)	(0.45)	(-1.96)
October	1.44	4.84***	1.37
	(0.23)	(3.19)	(1.19)
November	-2.00	9.86***	12.58***
	(-0.32)	(6.31)	(10.69)
December	-1.87	5.61***	-3.73***
	(-0.27)	(3.08)	(-2.72)
(Beginning of month)			
Middle of month	2.92	-0.80	-1.94***
	(1.01)	(-1.06)	(-3.37)
End of month	2.38	1.19	-1.44**
	(0.79)	(1.54)	(-2.43)
Constant	49.68***	103.40***	120.82***
	(7.94)	(50.50)	(77.22)
Year of listing	No	Yes	Yes
Number of Observations	4 353	142 316	275 304
R^2	0.034	0.031	0.035

6.3.2 SEARCH CRITERIA

I study the matched and unique samples to find differences around selected key search criteria; €5 000 below or above €100 000's in asking price and the same with 2m^2 more or less living space for flat tens of area. I begin with a hedonic regression and continue by deriving the desired coefficient signs for closer analysis. The selected variables and corresponding coefficients are listed in Table 13.

Table 13 Search criteria regression

In this table I list the results for the search criteria analysis. Figures in parentheses below the coefficients are the t-statistics and ***, **, * denote statistical significance at 1%, 5% and 10% levels respectively.

	Matched	Unique	Matched	Unique
Variable	Coefficient	Coefficient	Coefficient	Coefficient
Asking price dummy (€)				
95 000-99 999	-3.13	-24.02***	-9.17	-18.44***
	(-0.46	(-13.49)	(-1.40)	(-10.38)
100 000-105 000	4.39	-11.16***	5.76	-6.78**
	(0.36)	(-3.73)	(0.50)	(-2.29)
195 000-199 999	22.09**	2.01	26.40***	-1.42
	(2.51)	(0.98)	(3.15)	(-0.70)
200 000-205 000	29.08	67.37***	25.85	65.64***
	(1.45)	(15.02)	(1.37)	(14.79)
295 000-299 999	-12.09	17.65***	0.52	11.48***
	(-0.66)	(6.19)	(0.03)	(4.05)
300 000-305 000	125.34***	83.38***	121.03***	80.99***
	(3.11)	(11.07)	(3.18)	(10.86)
395 000-399 999	30.71***	29.85***	29.58***	23.51***
	(5.69)	(7.39)	(5.36)	(5.86)
400 000-405 000	-29.58	95.28***	-104.26	92.33***
	(-0.37)	(7.99)	(-1.21)	(7.82)
Area dummy (m^2)				
38-39.9	-13.05	-22.56***	-8.88	-5.81**
	(-1.15)	(-9.49)	(-0.76)	(-2.44)
40-42	-30.86***	-22.41	24.62**	-10.05***
	(-2.82)	(-10.22)	(-2.38)	(-4.58)
58-59.9	-6.81	-4.32**	-8.28	0.02
	(-0.97)	(-2.36)	(-1.23)	(0.01)
60-62	-1.89	-9.86***	-3.78	-6.79***
	(-0.23)	(-5.06)	(-0.49)	(-3.47)
78-78.9	8.46	5.31**	-3.63	-4.36**
	(0.91)	(2.50)	(-0.41)	(-2.02)
80-82	8.74	2.99	-1.38	-6.60***
	(1.03)	(1.46)	(-0.17)	(-3.18)
98-99.9	21.36	8.59***	9.30	-2.18
	(1.30)	(2.91)	(0.59)	(-0.74)
100-102	24.60	8.12***	7.96	-2.86
	(1.37)	(2.94)	(0.46)	(-1.03)
Constant	54.58***	84.44**	66.53***	53.60***
	(39.74	(233.75)	(9.50)	(61.66)
Zip code	No	No	Yes	Yes
Number of rooms	No	No	Yes	Yes
N	4 353	142 316	4 353	142 316
R^2	0.017	0.007	0.188	0.028

I put all the dummies in the same model and control it for zip code and number of rooms. As Table 13 shows I mainly get significant results for the unique sample whereas the low number of observations in the matched sample (see Table 14) results to lower statistical significances of coefficients.

To value my reasoning behind the search criteria I compare the results pairwise. The coefficient signs are not that important now, but rather the compared coefficient values. To support my idea the coefficients for asking prices below a flat €100 000 should be lower than for those just above. This would indicate a shorter TOM. Following the same logic with area, the coefficients for just below a flat $10m^2$ should be larger indicating a longer TOM as for those just above. This is because flat $10m^2$'s could usually be the lower limit of a search criterion and appear first in the search results due to lower price. On the other hand as the smaller apartments are in general cheaper in absolute terms, one could easier buy such for this reason only compared to somewhat larger one in her search limits. Table 14 shows the pairwise comparison results as well as amount of observations with the selected variables. These results clearly support my reasoning behind the importance of search criteria. The matched sample performs fairly similarly to the larger sample with unique observations.

Table 14 Search criteria results

In this table we see the results of the search criteria regressions based on my expectations. A plus sign indicates the difference between the coefficients follows my logic whereas a minus sign indicates the opposite. Signs in parentheses show results for the second regression controlling for zip code and number of rooms.

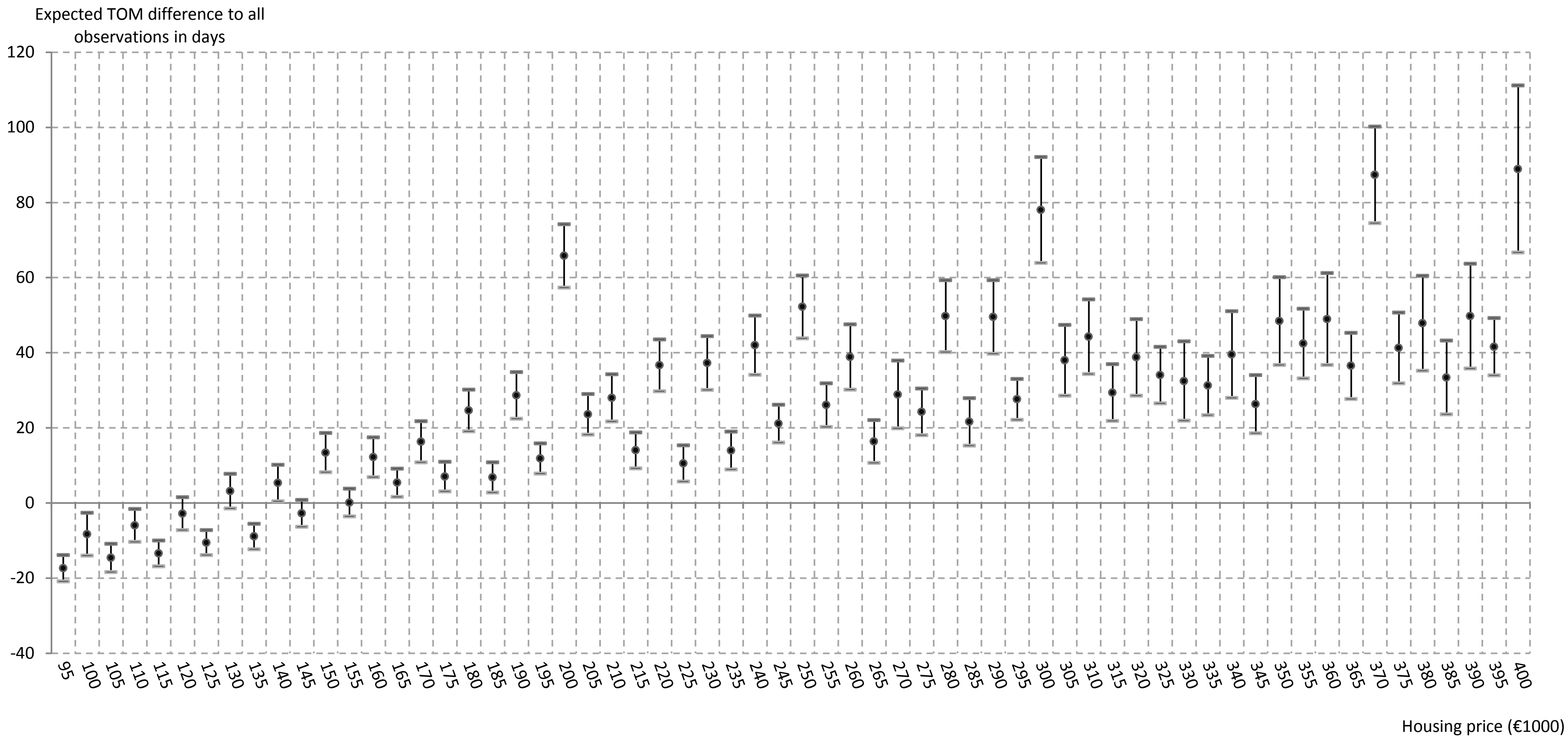
Variable	Matched Sign	Unique Sign	Matched Observations	Unique Observations
Asking price around (€)				
100 000	+ (+)	+ (+)	187	6 253
200 000	+ (-)	+ (+)	100	4 168
300 000	+ (+)	+ (+)	25	2 021
400 000	- (-)	+ (+)	259	977
Area around (m^2)				
40	+ (-)	- (+)	104	5 602
60	- (-)	+ (+)	235	8 244
80	- (-)	+ (+)	166	6 695
100	- (+)	+ (+)	44	3 542

I continue with these promising results and broaden the test to study all €5000 price ranges as dummies within the previous test limit of €95 000 to €405 000. I use the unique sample and control the test for zip code, construction year and year of listing. The results are presented with Figure 17. My first observation is that three out of four peaks in the figure are flat €100 000's thus included in my suggestions above. These prices stand out as the worst listing prices what comes to expected sale time. Although the €100 000 level does not stand out particularly, it clearly suggests a longer TOM than the previous or following dummy. Following this logic, somewhat surprisingly, I find that other than the €330 000 dummy, every category of below flat €10 000's suggest a shorter TOM than for the dummies surrounding with just above flat €10 000's. This means the seller should always, at least within my study limits, determine her list price below a flat €10 000 rather than just above. I find this result very interesting and I conclude that cosmic selection of listing price has a significant influence on expected sale time. One driver for this phenomenon could be that banks tend to offer mortgages with flat €10 000's and therefore the buyer may be lead to use the full amount, but not above it.

I also try the same procedure with the area dummies. Logically, the results are not as consistent as with the listing price, since presumably search criteria usually is at least 10 square meters or so. However, with flat ten square meters I find supporting evidence for six out of seven observations. A seller cannot, at least easily, change the area of her housing, but I find these results interesting when analyzing the buyer side behavior. According to my suggestion, if a buyer candidate searches for housing with 40 to 50 square meters of area, the smallest dwellings in this category seem favorable due to lower absolute price. I assume that when a buyer sets the area limits of her preferences, all dwellings are as good to her in terms of area. Thus, a similar priced dwelling with less living area should have other valuable features such as condition or locations that attract the buyer.

Figure 17 Effect of asking price selection to TOM

In this figure I plot results of the additional test to search criteria study. Each €5000 price range is represented by a dummy variable, first one being €95 000-€99 999. Tiny lines above and below observations represent the 95 percent confidence intervals. In this test I control for zip code, construction year and year of listing. N=142 316.



7 CONCLUSIONS

I end my thesis by concluding the study with key findings followed by discussion of possible future studies. As for the most individuals the transaction of a dwelling is the largest one of their lives, it is important to understand determinants of this trade. Other than the price and pricing of a dwelling, sale time is one key feature to estimate when pursuing a housing transaction. This is also a composer of housing liquidity that traditionally has not received very much of attention, neither among the academics nor the media. Media does publish average selling time data every now and then, but analysis behind the figure changes is often poor. Other than for the academics, my results should be appealing for any individual considering a housing transaction or interested in housing markets in general. For housing investors the results may be even more interesting, because their investor sentiment is likely to be lower, thus analyzing possible targets with more rational.

As discussed in the section 3.1, housing markets and housing as a commodity itself differs substantially from many other assets. As I earlier proposed, most people seem to take the market liquidity for granted without an analysis of why. Selling one's own home involves feelings such as devotion and sadness or happiness. Based on feelings the sellers may act greedy, busy or loss averse. Buyers can act irrational as well. When a buyer has found the ideal dwelling with desired attributes, she might pay excess price to get the asset quickly or generally to ensure the deal. All these behavioral aspects are assumingly stronger than with other assets and tend to lower housing liquidity as judgments are not based on asset value only.

I seek to reduce the information asymmetries surrounding the liquidity of housing and explain the variation in it by decomposing days on the market to various correlations with housing features and more. To my knowledge, I'm the first one to study the determinants of TOM in Finland. Globally, there is increasing interest towards real estate studying, and the Finnish housing market has been the focus of many novel studies as well. However, sale time or any modifications of it does not appear in a single paper I find. Hence, my aim is also to make this study diverse to serve basis for future studies.

The research questions I address with this paper are: How does marketing time of a dwelling depend on housing features, location and local macro-economic factors? Do submarkets exist and how they affect the marketing time in various Finnish housing markets? Are there behavioral patterns decreasing or fluctuating liquidity on the Finnish housing market? I use a data set collected from three suppliers totaling to some 400 000 housing selling advertisements. Post merging the data sets and removing observations with insufficient data I end up with 275 304 advertisement. I also form two subsamples: one with unique observations based on available data and another by merging the advertisements with actual transaction data for years 2011-2012.

The results show that based on housing and location features and the listing date, sellers may forecast expected selling time for their dwelling. Although this forecast does not presumably represent the absolute truth, it suggests what the selling time for a certain dwelling could be. The most value added from the first research question analysis, determinants of TOM in section 6.1, is definitely having a contribution to TOM for many housing and location features as well as macro-economic figures. These contributions are also measured in days for easy interpretation. In addition, most of the coefficients are statistically significant. Most of housing features yield expected results, but I find certain coefficients constantly surprising. As discussed in the results section, interpretations of these results are often multifaceted. For instance with price discount, you may think a reduction in price accelerates the trade, but on the other hand the seller may have tried to sell her dwelling for a long time before the price adjustment. Additionally, a discount may signal that the dwelling has not been able to attract other buyers either. In general, regressions on determinants of expected sale time yield 10-25 percent R^2 values. This means that 10-25 percent of variation with TOM is explained by the independent variables I test for. This comes as no surprise, as the housing markets and dwellings are heterogenic as earlier discussed. As a conclusion, it is easy to say that TOM is not fully predictable by any outstanding data, but coefficients still describe the Finnish housing market and a single housing sale time expectations fairly well.

Second research question deals with submarkets' effect to TOM. I divide Helsinki and Tampere markets into smaller submarkets based on official city practices and Statistics Finland categories. I study how TOM varies with different submarkets to each other and how these relations change over time. The results are two sided: I don't find highly significant variations between submarkets, but differences with the two cities and over time are visible from section 6.2. Pryce and Gibb (2006), who study the same method with Scottish data, find very interesting results in their study. Interpretation of consistency with my findings can be that the markets I study are more efficient or homogenous as the ones they study.

My final analyses, in section 6.3, deal with behavioral aspects on Finnish housing market. I determine correlations with TOM and listing date: weekday, month, time of month while controlling for year of listing. Results suggest that even with a careful selection of the listing weekday, the seller may expect more than a week of reduced time on market. With months the differences are even greater. Based on the analysis, the most favorable listing weekday has been Thursday. With months, the seller may expect shorter selling time when listing in late summer or around the change of year.

Lastly I study the effect on buyer's search criteria on expected sale time. I find that with careful price setting, even if only cosmetic, the seller may manipulate expected sale time by positioning her dwelling favorable in buyers' searches. A large portion of houses are currently searched via internet. When a buyer sets, say a price limit of her choice, she should find the more expensive housing more interesting in the group for larger area or other additional features. By capping the price, the dwellings just above this limit do not show up in the search and buyer might not even realize they exist. In other words, these may seem less favorable to the buyer or receive less attention. Results of this study are very interesting and follow my logic with all but one €5000 category within €95 000-€400 000 asking price range. This means with a price set just below a flat €10 000 limit, the seller should expect a shorter selling time than with any price just above the limit. I also find evidence of this phenomenon with square meters of area. The logic and results are opposite to price setting. For example, a dwelling with 40 square meters shows in a search criteria of 40 to 50 square meters. Other things equal, this dwelling should be cheaper than larger ones still supporting the buyer's preferences. On the other hand, a smaller and similar priced dwelling should have additional value features such as a better condition or location. Thus this dwelling is more attractive and sells quicker. In the analysis, I find positive results for six out of seven cases.

This study provides basis for future research by presenting results to compare with and to build new research questions on. With another time frame and new locations the study can already be very different from this one. Every added detail on the dwellings might give additional and interesting information. Due to the fact that expected selling time is not a formula anyone could come up with, I would be interested to learn more of the behavioral patterns on the market. Another great topic would be separating housing sold by a broker to individually sold and compare the results. Generally, an even more interesting study would analyze the effect of dynamic pricing on selling time or compare optimal time on market with selling price as e.g. Anglin and Rutherford (2003). However, regardless of the research question of interest, I emphasize the data quality.

I believe my study on search criteria effect on TOM could be expanded to other commodities as well. This might require a commodity of certain price level, but I see no problem applying the idea with cars for example. My findings suggest a behavioral pattern that probably not only relates to housing business, but could be generalized. Same kind of pricing can be found in grocery stores or actually in any consumer stores: large signs with “only €1.99” or similar written on them that sound better than a flat €2.00 for example. Other commodities, especially cars, could make a good topic for sale time determinants study as well. They are more homogenous than housing, but consumer behavior and consumption preferences clearly change over time. With more homogenous goods the explanatory power of tests should increase as well.

References

Anglin, P.M., 2006. Value and liquidity under changing market conditions. *Journal of Housing Economics* 15, Pages 293–304.

Anglin, P.M., Rutherford, R., 2003. The trade-off between the selling price of residential properties and time-on-the-market: The impact of price setting. *The Journal of Real Estate Finance and Economics* 26:1, Pages 95-111.

Arnott R., 1987. Economic Theory and Housing. In: Mills, E. (Ed.) Handbook of Regional and Urban Economics, Volume II. Elsevier Science Publishers.

Ashenfelter, O., Genesove, D., 1992. Testing for price anomalies in real estate auctions. *National Bureau of Economic Research*, Working paper No. 4036.

Bakos, J.,Y., 1997. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* , Volume 43, Number 12, Pages 1676-1692.

Bourassa, S.C., Hoesli, M., Peng, V.S., 2003. Do housing submarkets really matter? *Journal of Housing Economics*, Volume 12, Issue 1, Pages 12–28.

Boyle, M.A., Kiel, K.A., 2001. A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities. *Journal of Real Estate Literature*, Volume 9, Issue 2, Pages 117-144.

Deo, M., Srinivasan, K., Devanadhen, K., 2008. The Empirical Relationship between Stock Returns, TradingVolume and Volatility: Evidence from Select Asia-Pacific Stock Market. *European Journal of Economics, Finance and Administrative Sciences* 12, Pages 58-68.

DiPasquale, D., Wheaton, W.C., 1996. Urban Economics and Real Estate Markets. Prentice Hall: Englewood Cliffs, NJ.

Einiö, M., Kaustia, M., Puttonen, V., 2008. Price setting and the reluctance to realize losses in apartment markets. *Journal of Economic Psychology*, Volume 29, Issue 1, Pages 19-34.

Evans, M.D.D., Lyons, R.K., 2008. How is macro news transmitted to exchange rates? *Journal of Financial Economics*, Volume 88, Issue 1, Pages 26–50.

Firstenberg, P., Ross, S., Zisler, R., 1988. Real Estate: The Whole Story. *Journal of Portfolio Management* 14, Pages 22-34.

Friedman, H.C., 1971. Real Estate Investment and Portfolio Theory. *Journal of Financial and Quantitative Analysis.* Volume 6, Pages 861-874.

Genesove, D., Mayer, C., 2001. Loss aversion and seller behavior: evidence from the housing market. *Quarterly Journal of Economics*, 116, Pages 1233–1260.

Haurin, D., 1988. The duration of marketing time of residential housing. *Real Estate Economics*. Volume 16, Issue 4, Pages 396–410.

Huovari, J., Laakso, S., Luoto, J., Pekkala, S., 2002.Asuntomarkkinoiden alueellinen ennuste. *Pellervo Economic Research Institute*. Reports No. 185.

Huovari, J., Mäki-Fränti, P., Volk. R., 2006. Alueellisten asuntomarkkinoiden kehitys vuoteen 2009. *Pellervo Economic Research Institute*, Working Papers, No 86.

Juntto, A., 2007. Tulot ja kulutus 2007: Suomalaisten asumistoiveet ja mahdollisuudet. Tilastokeskus.

Kalbfleisch, J.D., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data. 2nd edition. New York: Wiley.

Kang, H.B., Gardner, M.J., 1989. Selling price and marketing time in the residential real estate market. *Journal of Real Estate Research*, Volume 4, Number 1.

Kaplan, E.L., Meier, P., 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, Volume 53, Number 282, Pages 457-481.

Kaustia, M., 2010. "Disposition effect", Chapter 10 in Behavioral Finance (Robert W. Kolb Series in Finance), H. Kent Baker and John R. Nofsinger, eds., John Wiley & Sons, Inc.

Klein, J.P., Moeschberger, M.L., 1997. Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer–Verlag.

Kluger, B.D., Miller, N.G., 1990. Measuring Residential Real Estate Liquidity. *Real Estate Economics.* Volume 18, Issue 2, Pages 145–159.

Krainer, J., 1999. Real Estate Liquidity. *FRBSF Economic Review*. Number 3.

KTI Kiinteistötieto Oy. KTI index results for year 2011.

KTI Kiinteistötieto Oy. Markkinakatsaus, Kevät 2012.

Laakso, S., 1997. Urban housing prices and the demand for housing characteristics: a study on housing prices and willingness to pay for housing characteristics and local public goods in Helsinki metropolitan area, *ETLA – The Research Institute of the Finnish Economy*.

MacKinlay, A.C., 1997. Event Studies in Economics and Finance. *Journal of Economic Literature,* Volume 35, Number 1.

Malpezzi, S., 2003. Hedonic pricing models: a selective and applied review. *Housing Economics: Essays in Honor of Duncan Maclennan,* Edited by K. Gibb and A. O'Sullivan. Blackwell.

Markowitz, H., 1952. Portfolio selection. *Journal of Finance*, Volume 7, Issue 1, Pages 77-91.

Miller, N.G., 1978. Time on market and selling price. *AREAU Journal.* Volume 6, Issue 2, Pages 164-174.

Moilanen, I., Terho, T., 2010. Cross-sectional variation of rental yields in the housing market. *Master's thesis, Aalto University School of Economics.* Helsinki.

Morawski, J., 2008. Investment decisions on illiquid assets: a search theoretical approach to real estate liquidity. Wiesbaden: Betriebswirtschaftlicher Verlag Gabler.

Mortensen, D.T, 1986. Job search and labor market analysis. In: Ashenfelter, O.C., Layard, R. (Eds.), Handbook of Labor Economics, Volume 2. North-Holland, New York.

Nikola, N., 2011. The effect of pipe repairs on housing prices. *Master's thesis, Aalto University School of Economics.* Helsinki.

Nyberg P., 2009. Essays on Risk and Return. Doctoral thesis, Hanken School of Economics, Department of Finance and Statistics, Finance.

Oikarinen, E., 2007. Studies on housing price dynamics. Doctoral thesis, Series A-9:2007, Turku School of Economics.

Pryce, G., Gibb, K., 2006. Submarket Dynamics of Time to Sale. *Real Estate Economics*. Volume 34, Issue 3, Pages 377–415.

Redman, A.L., Manakyan, H., Liano, K., 1997. Real investment trusts and calendar anomalies. *The Journal of Real Estate Research* 14, Pages 19-28.

Ritter, J.R., 2003. Behavioral finance. *Pacific-Basin Finance Journal*, Volume 11, Pages 429–437.

Shleifer, A., 2000. Inefficient Markets: An Introduction to Behavioral Finance, Oxford University Press.

Sirmans, G. S., Macpherson, D. and Zietz, E., 2005. The Composition of Hedonic Pricing Models, *Journal of Real Estate Literature*, Volume 13 Issue 1, Pages 3-43.

Springer, T.M., 1996. Single-family housing transactions: Seller motivations, price, and marketing time. *The Journal of Real Estate Finance and Economics*. Volume 13, Number 3, Pages 237-254.

Tyvimaa, T., Kananen, J., 2011. Suomalaisten asunnon hankinta ja asumisviihtyvyys, Asumista ja hyvinvointia tukevat alueelliset palvelumallit -hankkeen asukaskyselyn (2011) tuloksia. Tampereen teknillinen yliopisto, Rakennustekniikan laitos.

Uusitalo, O., 2008. Internet asunnon myyjän tukena. *Master's Thesis, Tampere University of Technology.* Tampere

Viitanen, K., Palmu, J., Kasso, M., Hakkarainen, Erja., Falkenbach, H., 2003. Real Estate in Finland. *Helsinki University of Technology, Kiinteistöopin ja talousoikeuden julkaisuja,* B 107.

Wood, J.H., Wood, N.L., 1985. Financial markets. Harcourt, Brace, Jovanovich.

APPENDIX

Table 15 Correlation matrix for the matched sample

Matched sample	TOM	Apartment house	Row house	Other type	Area	Ln(debt-free asking price)	Ln(debt-free trasaction)	Sauna	Balcony	Elevator	Asking price less transact	Built 1870-1949	Built 1950-1989	Built 1990-2012	Helsinki	Jyvskylä	Tampere	Turku	Kuopio	Oulu
Apartment house	-0.014	1																		
Row house	0.004	-0.865	1																	
Other type	0.022	-0.443	-0.067	1																
Area	0.154	-0.397	0.335	0.191	1															
Ln(debt-free asking price)	0.089	-0.190	0.144	0.120	0.583	1														
Ln(debt-free trasaction price)	0.073	-0.177	0.130	0.120	0.573	0.960	1													
Sauna	0.026	-0.245	0.203	0.124	0.138	0.094	0.071	1												
Balcony	-0.070	0.176	-0.143	-0.094	0.106	-0.090	-0.087	0.053	1											
Elevator	-0.076	0.355	-0.307	-0.158	-0.065	0.021	0.017	-0.017	0.297	1										
Asking price less transaction price (%)	0.057	-0.054	0.063	-0.005	0.072	0.193	-0.083	0.089	-0.001	0.018	1									
Built 1870-1949	-0.042	0.134	-0.146	-0.005	-0.063	0.294	0.305	-0.085	-0.376	0.045	-0.024	1								
Built 1950-1989	-0.151	0.065	-0.022	-0.091	-0.076	-0.416	-0.379	-0.134	0.285	-0.016	-0.131	-0.562	1							
Built 1990-2012	0.188	-0.216	0.171	0.124	0.163	0.220	0.180	0.251	0.029	0.008	0.136	-0.201	-0.652	1						
Helsinki	-0.115	0.133	-0.149	0.003	-0.064	0.504	0.524	-0.079	-0.109	-0.035	-0.025	0.286	-0.093	-0.140	1					
Jyvskylä	0.009	-0.077	0.097	-0.020	-0.012	-0.183	-0.189	0.045	0.021	-0.004	0.009	-0.109	0.025	0.065	-0.274	1				
Tampere	0.095	-0.030	0.020	0.023	-0.008	-0.128	-0.132	0.035	0.002	0.002	0.002	-0.063	-0.053	0.129	-0.420	-0.145	1			
Turku	0.020	-0.011	0.015	-0.005	0.095	-0.174	-0.194	-0.023	0.093	0.092	0.052	-0.085	0.108	-0.057	-0.350	-0.121	-0.185	1		
Kuopio	-0.002	-0.093	0.126	-0.040	0.046	-0.131	-0.121	0.045	0.069	-0.058	-0.042	-0.116	0.075	0.005	-0.278	-0.096	-0.147	-0.122	1	
Oulu	0.048	-0.011	-0.006	0.032	-0.030	-0.212	-0.222	0.032	-0.015	0.011	0.015	-0.100	0.005	0.077	-0.241	-0.083	-0.127	-0.106	-0.084	1