Aalto University
School of Economics

Cross-sectional Variation of Rental Yields in the Housing Market - Evidence from Finland

Finance
Master's thesis
Iikka Moilanen
2010

Department of Accounting and Finance
Aalto University
School of Economics

Aalto University
School of Economics

Cross-sectional Variation of Rental Yields in the Housing Market

Evidence from Finland

Master's Thesis
Iikka Moilanen & Tomi Terho
Spring 2010
Finance

Approved in the Department of Accounting and Finance __ / __ 2010 and awarded the grade ____

Aalto University School of Economics

Aalto University School of Economics
Master's Thesis
Iikka Moilanen & Tomi Terho (in alphabetical order)

Abstract
June 15, 2010

CROSS SECTIONAL VARIATION OF RENTAL YIELDS IN THE HOUSING MARKET – EVIDENCE FROM FINLAND

PURPOSE OF THE STUDY

The purpose of this thesis is to investigate whether rental yields in the housing market exhibit cross sectional variation, and if such variation does exist, what is causing it. Our hypothesis is that net rental yield declines as asset prices increase, a violation of the no arbitrage condition and thus, of the efficient markets hypothesis. Though this phenomenon has previously been documented, it has not been analyzed from an investments perspective. Therefore, we attempt to open up discussion on the causes and implications of cross-sectional variation of rental yields.

DATA

The main data in this study comprises of the Oikotie housing advertisement database, a large Finnish online home brokerage. The dataset includes all the data that advertisers enter when advertising a dwelling for sale or for rent. The data set covers a time period that begins in January 2002 and ends in September 2009 and after a process of matching advertisements of dwelling for rent to for sale advertisements, includes altogether 31,864 dwellings or dwelling pairs with data on both asking rent and sales price as well as several other dwelling-specific characteristics. Additional data has been gathered from the Statistics Finland database, the Helsinki Urban Facts database and the Google Maps service.

RESULTS

The results indicate that cross-sectional variation of rental yield exists and that it is strongly negative relationship with asset value. Thus it seems dwelling prices and rents are not considered simultaneously with a set rental yield in mind, but rather through separate processes. The difference in yields is economically significant, as the mean yield of the highest and lowest decile differ by up to 4 %.

In our regression analyses of net rental yields, we find statistically significant coefficients for variables reflecting dwelling size, age, type, location and many other factors, while controlling for time and macro-level factors such as unemployment and average income in the area. Although most of these coefficients are intuitive when considering rents and prices separately, the fact that net rental yield varies so greatly within a single market means that opportunities for profitable investment strategies exist, and are accessible even by using rather simple rules of thumb.

KEYWORDS

Rental yield, housing market, cross-sectional variation

Aalto-yliopiston kauppakorkeakoulu Tiivistelmä
Pro Gradu tutkielma 15.6.2010
Iikka Moilanen & Tomi Terho (aakkosjärjestyksessä)

CROSS SECTIONAL VARIATION OF RENTAL YIELDS IN THE HOUSING MARKET – EVIDENCE FROM FINLAND

TAVOITTEET

Tämän tutkielman tavoitteena on tutkia vaihtelevatko vuokratuotot asuntomarkkinan sisällä, ja jos tälläistä vaihtelua esiintyy, mitkä tekijät sitä aiheuttavat. Hypoteesinamme on, että nettovuokratuotto on matalampi kalliimmille asunnoille, mikä on vastoin tehokkaiden markkinoiden olettamaa sen heikossakin muodossa. Vaikka tuottojen vaihtelu on havaittu aiemmissakin tutkimuksissa, sitä ei ole aikaisemmin analysoitu sijoitusperspektiivistä. Pyrimme tällä tutkielmalla avaamaan keskustelua vuokratuottojen vaihtelusta asuntomarkkinoiden sisällä, sekä tämän vaihtelun syistä ja seuraamuksista.

AINEISTO

Keskeisin tutkimuksessa käytetty aineisto on Oikotie-internetportaalin asuntojen myynti- ja vuokrailmoitukset. Oikotie on suuri suomalainen asuntojen välityspalvelu Internetissä. Aineistoon kuuluu kaikki tiedot, jotka ilmoittaja syöttää palveluun ilmoitusta jättäessään, ja aineisto kattaa ajanjakson tammikuu 2002 – syyskuu 2009. Tästä aineistosta on etsitty ja yhteensovitettu yhteensä 31 864 ilmoitusparia, joissa vuokra- ja myynti-ilmoituksen kohde on ollut joko sama asunto tai toisiaan hyvin läheisesti vastaavat asunnot. Oikotien lisäksi aineistoa on kerätty Tilastokeskuksen sekä Helsingin kaupungin tietokeskuksen tietokannoista sekä Google Kartat –palvelusta.

TULOKSET

Tulokset osoittavat, että asuntomarkkinoiden sisällä esiintyy vuokratuottojen vaihtelua ja, että vuokratuotto laskee voimakkaasti kalliimpien asuntojen kohdalla. Tulokset viittaavat siihen, että asuntojen pyyntivuokria ja –hintoja ei harkita yhtä aikaa sopiva vuokratuotto huomioiden, vaan että molemmille on erilliset hinnoittelumenetelmänsä. Tuottoero on taloudellisesti merkittävä: ylimmän ja alimman hintakymmenyksen välinen vuokratuottoero on jopa 4 %.

Vuokratuotolle tehdyn regressioanalyysin avulla havaitaan, että asunnon koko, ikä, talotyyppi, sijainti ja moni muu tekijä ovat tilastollisesti merkitseviä vuokratuoton määrittäjiä. Analyysissä kontrolloidaan lisäksi ilmoituksen ajankohta sekä useita alue-tason muuttujia, kuten alueen työttömyys ja keskitulot. Vaikka useat näistä tekijöistä onkin helppo ymmärtää, kun vuokraa ja hintaa käsitellään toisistaan erillisinä muuttujina, vuokratuottojen voimakas vaihtelu alueellisesti suppeankin markkinan puitteissa tarkoittaa, että kannattavia sijoitusstrategioita on mahdollista löytää, vieläpä hyvin yksinkertaisia nyrkkisääntöjä käyttämällä.

AVAINSANAT

Vuokratuotto, asuntomarkkinat, poikkileikkaustutkimus

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Housing makes up a large part of global investment assets (66 % of Finnish investment assets and nearly two thirds in the US). Despite this, the field has received only limited attention from academic researchers. Traditional finance research focuses very strongly on securities markets where the availability of data is good and events are plentiful. Because housing assets, in addition to being investment assets, are also consumer durables, comparisons such as the one done above may not be completely relevant, but developing a better understanding of the investment landscape in housing markets is necessary to be able to understand the investment decisions faced by the majority of individuals. Therefore, we turn our attention to investing in the housing market.

Early equilibrium models of house prices have focused on a spatial equilibrium, pricing houses relative to their location from, e.g., the city center, services, central business districts, etc. (Muth 1961). These models are, however very imprecise, as the amount of factors to be considered is overwhelming. After spatial equilibrium, hedonic models emerged as the more developed versions of these purely spatial models (Lancaster 1966) where structural, neighborhood, location, contractual and temporal characteristics are used to determine house values. Still, the problems related to measurability of the factors remained.

Much of academic research in housing (see, e.g., Case and Shiller (1987, 1989)) uses a typical financial no-arbitrage condition, which stems from the capital asset pricing model. According to this condition, investors earn equal risk-adjusted returns from investing in the housing market or other markets. In this approach, housing is expected to return a flow of housing services, and its value can be calculated by discounting. However, one of its major problems concerns the issue of risk aversion. As stated earlier, a typical home owner's investment portfolio is quite undiversified, therefore it is difficult to assess whether the return is appropriate without knowledge of the individual's risk premium on housing.

Apart from the CAPM-based approach used by Case and Shiller, there exists another financial no-arbitrage condition, which states that there are no predictable excess returns to being an owner relative to being a renter (Poterba 1984). The rent-own decision is perceived mostly as a financial consideration, since the same level of housing services is received whether the housing unit is owned or rented. The main problem with this approach has been to ensure the comparability of the owned and rented housing, because rental units generally differ in terms of size, location, building type and other possible factors. Unbiased data sets that are free of the quality-adjustment problems have been rarely obtained.

This paper focuses on rent-own financial equilibrium and provides a fresh set of quality-matched data for analysis. Past studies have focused on evaluating undervaluation or overvaluation of the whole market on the basis of aggregate rent and price levels. (see. e.g. Oikarinen, 2007) They claim that it is sensible to assume the rent-price ratio to vary between regions and metropolitan areas, and consider different expectations of price appreciation to be the main explaining factor for the variation. However, we feel it is worthwhile to delve further into this variation and analyze it in the context of differently priced and sized housing within the same metropolitan area.

With new data of over 32,000 matched pairs of advertisements of owned and rented housing units, we attempt to overcome the aforementioned problem of comparability between rental and owned housing. By limiting arbitrage relationships to apply within the housing asset class, we attempt to neutralize the problem of risk aversion in terms of owning an undiversified portfolio.

Our primary hypothesis is that even after ensuring comparability between the rented and owned housing units, the rent-price relationship will exhibit significant variation across the cross-section of housing and that it will have a negative relationship with ask price, our proxy for asset value. Secondly, we do not believe that differences in expected price appreciation can explain the variation, but that the phenomenon has other explanations as well, which mostly relate to variation in the structure of supply and demand as well as cost of capital.

Earlier research of this phenomenon, other than on an aggregated scale, is very limited. Very recently, Garner & Verbrugge (2009), Tian (2008) and Hargreaves (2005) have began to examine the cross-sectional variance in the user costs of housing and rents and all have documented rental yields falling with price. Because these papers are written by scholars in the field of urban economics, their approach largely ignores analysis of whether the existence of the variation is consistent with the financial no-arbitrage relationship between renting and owning. From a financial economics perspective, it is a violation of the efficient market hypothesis if some parts of the market are more lucrative to own and in some parts it makes more sense to rent. We believe it is possible to examine this further, especially as the Finnish housing market and our data allow for a unique approach. We take the financial economics perspective in an attempt to invite further discussion in the field of arbitrage in the housing markets.

Due to our financial economics approach, our results also have implications for housing market participants. Typical Finnish households have strong preference to invest in their own home and become owner-occupiers, and possibly neglect the financial considerations. Our study aims to offer insight for the tenure choice decision-making. Moreover, the same results are applicable for any housing investor aiming to optimize the return and risk on the investment. From a diversified investors' perspective, the efficient market hypothesis is violated if one can engage in profitable asset-picking instead of passive investment strategies. For most investors in the Finnish market, cross-sectional variation in the residential rental yields is clearly an unchartered territory.

Our primary conclusion is that there indeed exists a discrepancy between rents and prices. It seems that dwellings are often not priced according to their cash flow producing capability, that is, net rental yield, but rather, that the appropriate ask rent and price are set through separate processes. As hypothesized, net rental yield is a function of dwelling value, which falls as value increases. The phenomenon is economically significant as we find that the yields of the highest and lowest priced asset deciles differ by up to 4 %. The phenomenon holds even after controlling for dwelling types, size, geography and date as well as many other factors in our regression analysis, discussed later. As hypothesized, the elasiticities of price and rent to dwelling-specific characteristics in the hedonic pricing model differ, which causes the variation, but clearly violates the no arbitrage condition and thus, efficient markets hypothesis.

This paper is structured as follows: the next section will provide background to the subject as well as broadly review relevant prior literature, theories and assumptions made as well as briefly describe the Finnish housing market. Section 3 will start by summarizing prior literature regarding methodology in hedonic models after which we will proceed to choose the functional form to be used in the quantitative analysis of this paper, as well as formulate our hypotheses. In section 4 we will describe the data used in this study as well as discuss some possible limitations that using the advertisement dataset imposes. In section 5 we will present the results of the regression analyses as well as further describe the data with several figure. Section 6 will provide a discussion of the results, our conclusions, their contribution and some suggestions for further research that can we done with similar approach data-wise.

2 Background

In this section, we broadly review the theory and assumptions behind the no arbitrage condition (NAC) model. We start by describing the dynamics of the housing market in a four-quadrant model and then proceed to discuss the financial no arbitrage condition, which is one of the four key equilibriums in the market. Next we go through assumptions of the NAC model and related literature. We then move on to the determinants of housing prices, rents and net rental yield, and then discuss the efficiency of the housing market. The section closes with a brief description of the Finnish housing market and Finnish households' preferences regarding housing consumption.

2.1 Four equilibriums of the real estate market

Since real estate is a durable good, its production and price are determined in an asset (or capital) market. In the real estate asset market, demand must equal supply. In the long run, the asset market should equate market prices with replacement costs that include building costs and land value. However, in the short run, the two may diverge significantly because of lags and delays that are inherent in the construction process. Generally, there are other determinants of asset demand besides the replacement cost of the assets. The most important determinant, as diPasquale and Wheaton (1996) point out, is the rental income that real estate assets earn. To understand rent, it is necessary to consider the market for the use of real estate, which can be referred to as the property market.

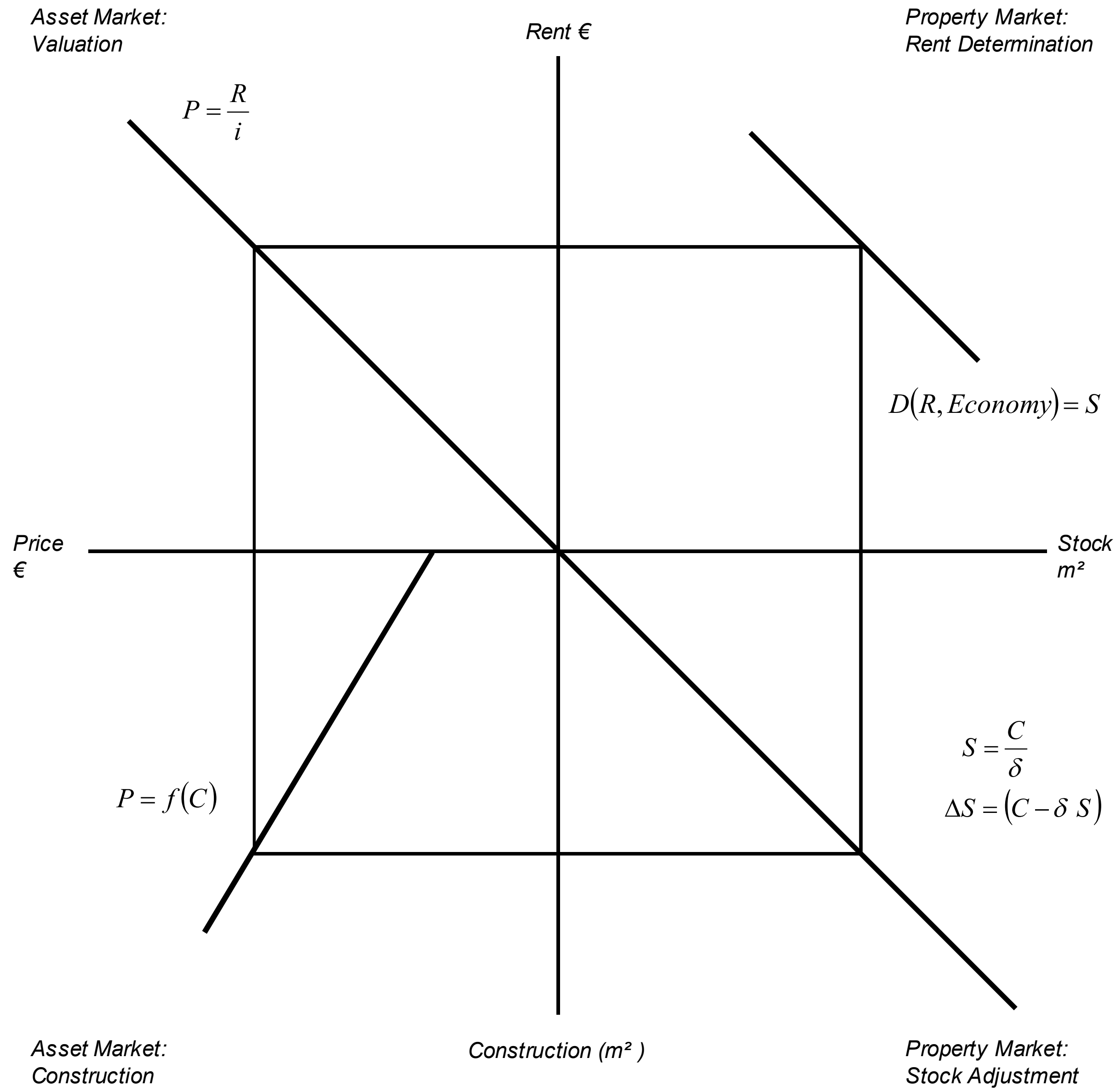
In the market for the use of real estate space, demand is formed by occupiers of space. In the residential market, which is the focus of this study, the occupiers of space are households that divide their income between the consumption of many commodities, one of which is space. Households are either tenants or owner-occupiers of residential space. For tenants, rent is simply specified in the rental agreement, and for owners, rent is defined as the annualized cost associated with ownership of property, i.e., the imputed rent or user cost.

The asset market and the property market are linked by two junctions. Firstly, rent levels determined in the property market are central in determining the demand for real estate assets. After all, investors are really purchasing an income stream when they acquire an asset. Thus, changes in rent in the property market immediately affect the demand for ownership in the asset market. The second link between the two markets is the construction or development sector. If construction increases and the supply of assets grows, not only are prices driven down in the asset market, but rents decline in the property market as well. These connections between the two markets are illustrated in the four-quadrant diagram adapted from DiPasquale and Wheaton (1996) in Figure 1.

Figure 1:

The four quadrant model

The four quadrant model (diPasquale & Wheaton, 1996) represents how equilibriums are formed in the housing market on both the asset market and the property market. Our study focuses on the north-west quadrant of the model.



The southwest and southeast quadrants of Figure 1 represent how the asset and property markets deal with changes in construction. The northeast quadrant has two axes, rent (per unit of space) and the stock of space. The curve represents how the demand for space depends on rents, given the state of the economy. Inelastic demand for households would be depicted by a vertical curve and demand sensitive to rents a more horizontal curve. An upward or downward shift occurs through an increase in households or economic growth.

This study is mostly interested in what is happening in the northwest quadrant. The first part of the asset market has two axes: rent and price. The ray from the origin represents the capitalization rate for real estate assets. This ratio is also known as the ratio of rent-to-price (or sometimes inversely as price-to-rent). This is the current yield that investors demand in order to hold real estate assets. Generally, four considerations make up this capitalization rate: the long-term interest rate in the economy, the expected growth in rents and prices, the risks associated with the rental income stream and price growth, and the treatment of real estate in the federal tax code. In the long run and considering a CAPM-type risk premium for the whole asset class, the capitalization rate can be assumed to be an exogenous variable used to determine a price P for rental level R. However, in the short run there are deviations from the long-run equilibrium capitalization rate. We will discuss the determinants of the capitalization rate (or rental yield) in the following sections.

2.2 The basic financial no-arbitrage condition model

The basic model of housing values, as proposed by Mills (1972), assumes that the housing market is perfectly competitive, which drives profits to be zero. Alternatively, gross returns to housing capital (income or implicit income) equal costs. This means that gross housing rents equal total costs of home ownership, which forms the basis of the NAC model, which will be discussed in this section.

Total costs of home ownership include both operating costs and the expected capital gains, which are stated in the following equation:

$$R = iP + mP + t_p P - gP, \tag{1}$$

which can be rearranged to obtain

$$R = \left(i + m + t_p - g\right)P, \tag{2}$$

where rent R is equal to the costs of ownership, which are an annual interest on the investment at rate i, expected capital gains at rate g, operating costs (maintenance, insurance, utilities) m and property taxes t_p.

For a house value of P and the costs implied by i, m, t_p and g, R is the rent that must be charged to break even (earn a normal return on capital), either as a landlord or as an owner-occupier choosing between renting and owning. This implies that consumers of housing should be indifferent between home owning and renting.

Suppose that the real interest rate i and the real growth g in market values are independent of the rate of inflation, π. This concept can be represented by replacing i by $(r + \pi)$ and g by $(g_r + \pi)$ in equation (2).

Therefore, equation (2) can be rewritten in real euro terms as

$$R = (r + \pi)P + mP + t_pP - (g_r + \pi)P, \tag{3}$$

$$R = rP + mP + t_pP - g_rP, \tag{4}$$

where the cost of capital is defined in terms of the real interest rate, r and the housing market appreciation is defined in terms of real growth in market values g_r.

The first term in equation (4) is the cost of foregone real interest that the homeowner could have earned through investing in another asset. This one-year cost is calculated as the price of housing times the long-term real interest rate, r. The second term is the one-year cost of property taxes, calculated as the house price times the property tax rate t_p. The third term m reflects operating costs expressed as a fraction of the house value, P. Finally, the fourth term is the house price times g_r, which is the expected real rate of house price appreciation. The sum of these four terms gives the total return to housing capital.

If the time period would span a long period, including periods of changing monetary policies and business cycles, the use of real growth rates would give better representation of market conditions. Since we are dealing with a relatively short time period and generally low and stable inflation rates within that time period, we deem the correction for real growth unnecessary. Therefore we return to equation (2).

Equation (2) can be modified by adding the tax deductibility of mortgage interest for owners who itemize on their capital gains or federal income taxes as follows:

$$R = \big((1-t)i + m + t_p - g\big)P. \tag{5}$$

The right hand side of Equation (5) is the annual cost of ownership, expressed in euro terms, and the equation in the brackets is the cost per euro of the total house value.

In the context of our study, Finland, property taxes and operating costs are allocated to individual dwellings through monthly maintenance fees to the housing cooperative[1]. Rearranging equation (5) we obtain

$$R_{net} = ((1-t)i - g)P, \tag{6}$$

where R_{net} is the net rent after deducting the operating costs and property taxes from the gross rent. This is the periodical explicit or implicit cashflow produced by the housing unit, either in the form of rent payment from tenant to landlord or as a rental equivalent yield for the owner-occupier of a housing unit. Some authors argue for the use of a constant depreciation rate (e.g. Poterba, 1991; Himmelberg et al., 2005), but we have not included such a variable as it somewhat unnecessary for the analysis of differences between cross-sectional capitalization rates. There is no consensus on how the depreciation rates should be attributed to different types of housing units. In the US, some authors propose higher depreciation rates for low-income housing projects due to tenant characteristics, but such arguments do not seem to be applicable to the Finnish market.

Dividing both sides by price P, we come up with the total return for any housing asset that are either explicit for the landlord or implicit for the owner-occupier:

$$\frac{R_{net}}{P} = (1-t)i - g. \tag{7}$$

It is intuitively clear, and evident from the equations above, that if rents decrease in value, it should trigger selling of assets held for rent, and buying if they increase. On the other hand,

---

[1] For simplicity reasons, we have assumed the maintenance fee and property tax of single-family houses (which are typically not reported in the for sale advertisements) to be at the same per square meter level as of detached housing cooperatives.

when considering properties used as a home, the same equation should hold: if the same standard of living is achievable by renting with a much lower cost, sale of the property becomes rational. Alternatively, it does not make sense to pay more rent than the same property's total cost of ownership would be when living in a rental home, so buying the property would be rational. Thus, prices and rents should continuously adjust to the long-term equilibrium level. In addition, for units with similar expected appreciation characteristics, the rental yields should essentially be the same and not dependable upon any other attributes, which is the focus of this study.

Naturally, an important simplifying assumption is made about interest rate, *I*; perfect capital markets are assumed. The cost of borrowing (mortgage rate) is assumed equal to the opportunity cost of capital (alternative investment rate for the consumer). On the other hand, the interest deductibility (*1-t*) depends on the loan-to-value (LTV) or leverage ratio. In Finland there are also euro term ceilings on the maximum tax deductions for income taxes, which will be further discussed in section 2.2.4.

Some authors have argued for the use of a separate risk premium exceeding the mortgage interest rate. Flavin and Yamashita (2002), Himmelberg et al. (2005) and Finicelli (2007) assume a risk premium of 2%. Poterba (1991) uses a 4% risk premium in user cost calculations, while Englund et al. (1995) McCarthy and Peach (2004) and Quigley and Raphael (2004) do not include an additional risk premium for owning housing assets. Discussion on risk can also be expanded to the uncertain future development of rents. Thus, price risk of owning housing assets and risk of rising rents would then be needed to be considered separately. Sinai and Souleles (2005) argue for ownership as hedging against future unanticipated movements in rental price. In their view, risk averse individuals "lock in" their future cashflows by buying into a desired housing mix instead of exposing themselves to potentially rising rents. Housing prices would thus capitalize a premium for avoiding net rent risk. In this study, however, we feel no need to separate risk premium from the expected appreciation unit, which will be discussed later in section 2.2.3.

Another concept that can be linked to both the housing unit depreciation and risk premium is the "land leverage" as introduced by Bostic et al. (2007). They argue that changes in the overall property value will depend critically on how much of its total value is contained in the land, a proportion they define as land leverage. As housing units are a bundle of physical structures and the land underneath the structures, it can be argued that only a proportion of the total asset value is depreciating over time. Thus, housing units with low land value would have higher depreciation rates than units with similar structures, but located in high land value areas. Investors would intuitively then require higher yields to compensate as the depreciation unit in the NAC equation would be higher. Simple arithmetic on somewhat extreme fluctuation in land value proportional to total value would signal differences up to 1% in net rental yield, assuming similar cost and depreciation for structures and land value ranging between 15 - 60% of total housing unit cost. However, Bostic et al. also show that housing assets with high land proportion are also more volatile and exposed to economic shocks in greater magnitude, which would likely offset differences.

Unfortunately, there are also several complications with the empirical application of the no-arbitrage condition. Due to high transaction costs and low liquidity of housing there is always slight divergence from the presented relation even if everyone in the market is fully rational. Credit constraints for the low-income deciles of the housing consumers certainly play a significant role, allowing landlords to benefit from the borrowing constraints of those who simply are unable to choose freely between taking a mortgage and renting. The transaction costs are also somewhat difficult to take into account, although with pre-planned housing tenure horizons one could make NPV calculations based on the no-arbitrage formula and include the transaction costs in the model. However, the investment horizon is immeasurable as well, and variation in the investment horizon could have an effect on the required rate of return through a risk component, which would vary in conjunction with risk aversion.

2.2.1 From housing NAC to the Gordon growth model

The no-arbitrage condition is closely related to the present value condition, i.e., to the fact that in an efficient market the price of a house must equal the present discounted value of its future net service (cash) flows. The two relations present the same idea somewhat differently. There is clear equivalence between the conventional Gordon growth model and the NAC. The Gordon growth model is presented as

$$P = \frac{C}{r - g} \tag{8}$$

where C is the first period cash flow from the asset, r is the required rate of return and g is the expected growth rate of the cash flows. From (8) it is easy to see that in the no-arbitrage relation C equals R_{net}, r is the cost of capital $(1-t)i$ and g stands for the expected appreciated rate. While in the conventional Gordon model it is the dividend and the rental growth rate that matter, in the NAC, the growth expectations are equal to the expected appreciation of housing in the next period. These can differ at times due to lags in adjustment of rents and prices, but in the long-run housing price growth and rental price growth are expected to be tightly related.

Looking back to the previous section, the left hand side of equation (7) is often referred to as the rent-to-price ratio or the capitalization (CAP) rate in real estate literature. Following the Gordon analogy, it is easy to understand this yardstick for relative valuation: the discount rate used to capitalize current cash flows. In this study CAP rate and rental yield are used side-by-side as they refer to the essentially the same concept. Different notations in literature arise from use of either gross or net yields, depending on whether maintenance costs, taxes or other operating costs have been deducted. Naturally, the use of micro data enables the use of net rental yield and net CAP rate, allowing for greater accuracy.

Following the intuition of the Gordon model and the financial equilibrium model built on it, this study investigates the effect of several factors that could potentially influence the capitalization rate through their respective effect in the growth and discount rates.

2.2.2 The model in literature and the contribution of this study

Since the market relationship between house price and rental price is an important concept in the analysis of housing markets, the basic model of housing values is used widely in the housing literature. Mills (1972) shows the model starting with a simple case, in which rent and house value are linked only by interest rate. In this example, homeowner's only cost is the annual interest. Later, they incorporate taxation, capital gains, inflation and operating costs proportional to house price into the model. If the rental market is competitive, annual rent just covers annual cost. In other words, annual rent equals the annual cost of home ownership. Poterba (1984) suggests that the marginal benefit (implicit rent) and cost (the user cost of the asset) of owning the housing assets are equal to homeowners. The user cost of the asset, defined originally by Jorgenson (1963), is the sum of the after-tax opportunity cost of holding the capital asset, after-tax property taxes, and depreciation and repair, less the expected capital gain of the asset. Poterba (1984) uses this model to argue that inflation reduces the effective cost of homeownership and raises the tax subsidy to owner occupation. He shows how changes in the expected inflation rate affect the real price of houses and the equilibrium size of housing capital stock. Himmelberg et al. (2005) use the model to study rent-to-price ratios in housing markets. They measure the annual cost of homeownership for 46 metropolitan areas in the US over 25 years and compare them with local rents to determine the changes in the level of house prices. They find that, from 1995 to 2004, the cost of ownership rose somewhat relative to the cost of renting. Also, they demonstrate that house prices are more sensitive to changes in real interest rates when rates are already low.

There have been many other papers employing the no arbitrage condition (e.g. McCarthy and Peach 2004, Girouard et al. 2006, Finicelli 2007) to assess if housing prices on aggregate level are misaligned in one or more countries or cities. The empirical application of the condition involves several problems, such as the evaluation of maintenance costs and comparability of rental and ownership data. Glaeser and Gyourko (2007) discuss this problem especially in the US context, where the owner-occupied versus rental stock of housing are physically very different, located in different parts of the metropolitan area and have access to a very different set of amenities. In our study, we attempt to avoid the metropolitan-level problem by collecting information about variables at the neighborhood-level in our sample of Helsinki Metropolitan Area.

However, the main problem with using the NAC model is the catch-all variable of expected appreciation, as it by its nature is the balancing force in the equilibrium – any consideration of over- or undervaluation is essentially a discussion on whether the appreciation rates implied by the model are sensible or not.

2.2.3 Key to the model – the catch-all expected appreciation variable

The expected appreciation variable of the model has received many different considerations in the studies focusing on variation of the aggregate equilibrium rental yield. Typically, simple and often somewhat arbitrary assumptions concerning the expected appreciation are used in the literature. Poterba (1992), Girouard et al (2006) and Garner and Verbrugge (2009) assume that expected housing price appreciation equals the expected rate in overall inflation. Englund et al. (1995), in turn, employ the average capital gain as the sum of long-term inflation expectations and historical growth in real rents. Himmelberg et al. (2005), who also include a forward-looking component by adding the spread between short- and long-term interest rates in the user cost formula, use the average real growth rate of housing prices from 1940 to 2000 as a proxy for expected real appreciation. Furthermore, Smith and Smith (2006), while building an NPV model to assess potential overvaluation, use the same arbitrary expected appreciation values (in the base case 3%) for a number of different areas in the US.

When analyzing the housing markets as a whole and focusing on the aggregate rent to price figures as a yardstick for over- or undervaluation, the measurement of the expected appreciation is of major importance, since different methodologies can lead to different conclusions about the extent of misalignment in housing prices. Obviously, the use of the purely backward-looking expectations utilized in several papers may induce misleading conclusions. Past price appreciation or overall inflation do not necessarily represent expectations. For example, if housing prices have risen rapidly during the past few years and are currently notably above the fundamental level, backward-looking expectations based on relatively short history would also imply fast housing inflation in the future. On the contrary, rational agents would be likely to take the prevailing overpricing into account. That is, rational agents would expect the adjustment of housing prices towards the fundamental level and, therefore, the forward-looking expectations would predict a substantially lower appreciation figure.

When applying the discussion above to the focus of this study, i.e. the variation of the net rental yield between different types of housing units instead of longitudinal aggregate variation, the question arises whether the potential variation of the net rental yields could simply be interpreted as variation in the expected appreciation rates - i.e. particular types of apartments in certain locations would be expected to gain more in value. Location clearly has major significance in the development of housing prices, and thus it is easy to interpret that some areas are expected to gain more value than others.

Capozza and Helsley (1990) as well as Capozza and Sick (1994) model variation of the expected appreciation by geographical regions. Their investigations of urban growth with uncertainty suggest variables such as income, income growth, population, population growth, developable area in a city, construction costs and lagged growth of population and income lead to geographical variation in expected appreciation rates. Usually these variables are employed to metropolitan level data, and intra-city variation is rarely considered. Studies, such as Himmelberger et al. (2005) and Chichernea et al. (2008), show the geographical variation in aggregate rental yields as a result of these demand and supply side constraints driving the expected appreciation in different metropolitan areas.

There are a very limited number of studies which would link values, rents and potential differences in appreciation rates within metropolitan areas and across city neighborhoods. Janssen, Soderberg and Zhou (2001) use hedonic models to explain capitalization rates across building types, ages and four specific locations in Stockholm. Rather than the issue of cross-sectional determinants of valuations, they are more concerned with estimation methods than interpreting the results for the variation in their observed capitalization rates.

Intuitively, there is no reason why the expected appreciation should vary between housing units with different characteristics other than location. Macro trends suggesting more demand for certain types of housing in the future are difficult, if not impossible, to foresee, and thus it is also difficult to believe the market as a whole would form rational expectations about the relative performance of, for instance, small studios versus large multi-room apartments in the long run. We will return to the preferences of Finnish housing consumers later in section 2.5.1.

2.2.4 The impact of inflation, tax breaks, tax subsidies and social housing benefits to owner-occupied housing

The impact of inflation on housing demand can be shown by using the basic housing value model. When inflation rates increase, nominal interest rates are augmented, and there will be an increase in both homeowner`s interest charges and nominal capital gains. However, an increase in the inflation rate reduces the real cost of homeownership because homeowners can deduct mortgage interest payments from their taxable income. This will lead homeowners to gain on balance. They will receive the full nominal capital gains, but they will pay only some of the interest charges. This suggests that there is a positive relationship between inflation and housing demand. Since the cost of ownership declines with a higher inflation rate, people are more willing to buy houses instead of renting, which results in an increase in real house prices. For example, Poterba (1984) shows that the high inflation rates in the 1970s led to a 30 percent increase in real house prices and the stock of owner-occupied housing. However, J.R. Kearl (1978) argues that the financial position of homeowners is not affected by the expectation of inflation in the long-run. Even though the debt of homeowners rises in the present, inflation may provide higher nominal incomes to compensate higher payments. Historically, this has been the case in the Finnish context, but it is nowadays limited to the individual’s ability to itemize interest payments in their capital gains taxation as ceilings are in place for deductions from income taxes (Oikarinen 2007).

So far, we have discussed the tax deductibility of mortgages and after-tax cost of capital, while ignoring the other two tax subsidies attributable to owner-occupied housing: the tax exemption of living in your own house and the tax exemption of capital gains from selling your own house.

Oikarinen (2009) emphasizes that while owner-occupants are not taxed for the implicit income produced by the housing assets in their possession, a capital gains tax is imposed on the rental returns (net of maintenance costs and interest payments on the loan that is borrowed to buy the dwelling) in Finland. Moreover, housing appreciation is taxed when a rental dwelling is sold. Hence, because of the tax benefits of owner-occupied housing, housing is worth less for investors than for owner-occupiers if the other variables in equation (1) are the same for both groups. Therefore, if housing is of equal worth to owner-occupants and portfolio investors, the required return for investors has to be set lower or the future expectations have to be more positive than those of the owner-occupants. The return required by a portfolio investor may be relatively smaller due to the greater diversification benefits gained by owning multiple dwellings (and possibly also other assets) in a portfolio. In addition, investors may be less risk-averse than households and may have lower interest costs on debt. Furthermore, professional skills and scale economies of a large investor may lower maintenance costs. However, without precise knowledge on the holding periods and capital structures, accurate estimates of tax effects on the equilibrium cannot be calculated.

Housing subsidies provided by the government could create segmented market spaces, especially in the low-income end, where individuals would prefer renting to owning. In Finland low-income individuals can claim housing allowance for up to 80% of housing costs[2]. The allowance is subject to multiple limitations on housing unit size, costs, and individuals' income and net worth. Other subgroups of market participants such as students and pensioners might distort the overall rent-setting if subsidies drive choices between renting and owning. Subsidies can clearly create market segments where renting is more desired than owning, but these lasting effects should influence prices as well: higher rents would be priced in as investors would compete for assets where natural clientele exists. Thus, any unequal treatment of owning and renting should have no effect on net rental yield on effective markets.

---

[2] Source: www.kela.fi (Kela – the social insurance institution of Finland)

2.3 Determinants of value, rent and rental yield

In the previous sections, we have concentrated on the formation of asset prices as a purely financial process, where relevant considerations are required rate of return, expected appreciation, risk and regulatory issues such as taxation. Due to the dual nature of housing as an asset class (investment as well as consumption), these considerations are not enough to gain a full understanding of the market. As will be discussed later, the housing market cannot be characterized as efficient, even of the weak form (see. e.g. Case and Shiller (1989)), and therefore, understanding the formation of supply and demand for properties on the consumer preference level is also necessary.

This section reviews several studies that form a basis for our hypotheses. Firstly, we go through perhaps the most puzzling observation about rental yield – the negative relationship between rental yield and house value. We then go through other studies that have addressed the issue of variation of capitalization rates, and finally, we turn to detailed hedonic studies that have deployed variables relevant to this study in the formation of hedonic house price or rent functions.

2.3.1 The relationship of rent and value

As discussed in section 2.2.3, previous literature has paid relatively little attention to the determinants of rental yield, merely interpreting the deviation as differences in expected appreciation. While focusing on the housing markets as a whole at metropolitan levels, the question of which particular housing assets are profitable to own has not attracted interest in the academic literature reviewed here.

Nevertheless, some evidence on the determinants of rental yield can be found from past literature. Most recently Garner and Verbrugge (2009) and later Tian (2008) have found that estimated rents are concave with estimated value, i.e. rents increase at a decreasing rate relative to value. In the absence of a credible explanation, this raises concern about the efficiency of the housing market. It is intuitively clear that there should be no relationship between value and return in real estate pricing models where CAP rate does not vary with value are standard and following EMH. Taking an analogy from the stock market, a share costing 10 Euros should have very similar return characteristics to a share costing 100 Euros, all else equal.

In the US, the reasons to study the rent-to-value relationship and other determinants of cap rate have been mainly statistical. The Bureau of Economic Analysis uses capitalization rates to incorporate housing services produced by residential housing assets to national income and product accounts. According to Tian (2008) they use a convex CAP rate (falling with value), but the deviations mostly occur at the very low end of housing values, which is explained by the use of gross income, ignoring depreciation and operating costs, which could be an unusually large proportion of the gross rent. In our study, these maintenance cost issues are avoided because we have proper data for the periodical maintenance fees including property taxes.

Garner and Verbrugge (2009) have used owners' estimates of rental equivalence and value reported in the US Consumer Expenditure survey to study the determinants of reported rental equivalence. They find that rental equivalence is concave in value. Since owner estimates of rental equivalence and value are not transaction prices, it is difficult to see why owners of higher value units should report significantly lower rent-to-value ratios, particularly when the question is asked in four successive interviews as part of successive waves of the Consumer Expenditure Survey.

Following this finding by Garner and Verbrugge, and in an attempt to explain it, Tian (2008) collects a dataset on 148 matched dwellings in Washington DC during a one-month time-frame. He concludes that the rent-value relationship is clearly concave. Tian's data consists of single family and multifamily housing units worth between 105,000 and 1,125,000 dollars. As Tian points out, the first possible explanation would be that there is possibly a case for higher required yields for very low value rentals arising from negative occupant characteristics including credit risk and higher depreciation, but that all of these effects should disappear well before the $105,000 value segment. It is worthwhile mentioning that in Finland landlords are typically covered for credit risk by a functioning guarantee rent system, where occupants deposit typically 1-3 months' rent in advance to a mutual bank account or simply pay similar advances to the landlord that are only returnable upon moving out. Thus, low-income housing units should not bear higher rental yields due to differing credit risk or other quality considerations regarding the renters.

Hargreaves (2005), reporting the results of a survey of residential landlords of 1,585 New Zealand rental properties, finds a similar decreasing relationship between value and rent. He also finds the functional form of the trend to be logarithmic rather than linear. Hargreaves' interpretation of the result is that renters are simply unwilling to pay for some additional features that add value to the house, such as extra land, views and extra bathrooms.

The second possible explanation to the rent-value puzzle would be different expected appreciation rates, meaning that more valuable housing units would gain more value going forward. Although this seems an attractive explanation, it is at odds with the basic principles of EMH and should not be considered reasonable.

As an alternative explanation, Tian (2008) proposes assumptions of a segmented housing market, where the number of renters falls simultaneously with the cost of renting i.e. the net rental yield. The model is based on the fundamental insight by Titman (1982) of individuals' tenure choice as a comparison between user costs comprising of mortgage payments and operating expenses and the market rent, as already discussed under our model assumptions. Tian's model proposes a segmented housing market, and it relies on the fact that user cost of households varies inversely with income due to the tax treatment of owner-occupied housing and the relatively lower borrowing cost facing households with higher income, as in the US, they can itemize more interest deductions[3]. Furthermore, because housing consumption rises with income (even for rental housing) and also varies inversely with user cost, housing value should increase with income. In their studies focusing on the taxation aspects of the issue, Palmon and Smith (1998) found a negative correlation between user cost and physical size of housing units for single family housing units in Houston.

Following from the insight that user cost in the US falls with the size of the mortgage (and house value), the high-end homeowners' cost measures are permanently lower than the professional landlords with constant capitalization rates. This, combined with absolutely low level of rental demand in the high end, creates a situation where equilibrium does not exist. The market is then cleared by "inframarginal" landlords who have a number of idiosyncratic reasons to retain the ownership of their housing unit (for example due to personal attachment to a particular unit, a desire to avoid a taxable event associated with sale of the unit etc.) and are therefore satisfied with a lower return. This, in turn, forces the professional landlords out of the high-end of the market. In the low-end of the market, where demand for rental

---

[3] In the US system, there is no upper limit on one's ability to deduct interest payments from taxable income. Thus, the ability to "save" on tax deductions increases with the marginal tax rate.

apartments is higher, these inframarginal landlords earn similar yields to the professional landlords, whose elastic supply clears the market at the equilibrium capitalization rate.

Tian's argumentation is intriguing but requires more inspection when applied into Finnish context. The proposition of falling user costs does not apply as the tax deductions are made from capital gains which are taxed at the flat 28% rate. In addition, the user cost is actually rising with the value of the home for the individuals who take benefit of the capital income deficit crediting, which allows them to deduct the mortgage payments from their income taxes. As these deductions have maximum amounts imposed, the user cost is rising after the ceiling is met.

The more appealing argument is the proposition of inframarginal landlords. As we will discuss later in chapter 2.5.1, Finns have strong preferences to move towards ownership tenure as income rises. According to Juntto (2007) ownership dominates the desired tenure choice already from the 3rd income decile onwards with the proportion of renters falling by every decile and only 1% of households in the top income decile report renting as their most desired tenure option. Combining this with the inframarginal landlord proposition, one could propose that the amount of landlords wishing to rent their housing units at rates below equilibrium levels could well be enough to cover the whole demand in the housing market. In this case the conclusion would be that Finns disregard financial considerations and opt for ownership even if there would be cheaper housing options available in the rental markets. This would inflate the prices compared to rents and allow for such disequilibrium to exist.

2.3.2 Other studies on rental yield determinants

Hundreds of studies have documented various specifications of hedonic models with either house value or rent as the dependent variable. However, as discussed earlier, due to severe limitations in availability of data, there have been few attempts to explain the relationship between these determinants and their joint effect on the rental yield. A common approach is to combine price series data of rents and values, but this is usually seen as problematic due to the potential bias arising from the possible differences in the attributes of the two series, i.e. comparing apples to oranges (see also section 2.2.2). In the following, we review few of the rare previous studies that have made an attempt to overcome these data limitations.

Janssen et al. (2001) study capitalization rates with a data set of 351 multi-unit properties in Stockholm. Their variables only include age and 3 dummy coefficients for locations (East, South, and West) within Stockholm. They find high consistence for both rents and values for whole apartment buildings in the inner city of Stockholm, explaining 97 and 85 percent of the respective variation in the logarithmic rent and value functions. Increasing age seems to lower both rents and prices, but the age effect is stronger with respect to prices, driving up the capitalization rates for older apartment buildings. As they have no information on the maintenance costs, they conclude that older buildings probably cost more to maintain and have a shorter remaining life. However, they note that the option to redevelop these buildings should be thought to have some value. One particular area, Stockholm East, offers lower rental yield because prices are higher, but rents compare with other parts of the city. Janssen et al. assume that this is due to more stable tenants, lower maintenance costs and less risk than in other areas.

Hattapoglu (2009) studies two micro datasets from Houston metropolitan area using house price and rent listings between 1996 and 2007 from Houston Association of realtors. In his study, Hattapoglu matches rents and values by determining separate hedonic rent and price functions for every neighborhood in the Houston area. He then compares the implicit rents to actual prices, and vice versa, with the model similar to the one employed in this study. He finds that size in square feet, distance to the central business unit and past price appreciation have a negative relationship with the estimated rental yield, while the number of bedroom has a positive coefficient.

Hattapoglu does not include value as a predictor for the rent to price ratio. He interprets the negative coefficient of size to result from higher income consumers purchasing larger apartments, and he claims that lower capitalization rates result from lower mortgage fees for high income consumers. Moreover, the positive coefficient for the number of bedrooms may indicate that consumers prefer houses with fewer bedrooms that have more per square footage relative to houses with more bedrooms that have less per square footage. These findings are interesting for our study, as we have the possibility to explore these effects even further with a more complete dataset, whereas Hattapoglu focuses only on houses and neglects some potentially interesting variables.

Importantly, Hattapoglu shows that the current rental yield is dependent on past price appreciation, but not on past rent appreciation. Thus, Hattapoglu concludes that people’s appreciation expectations are based on past price appreciation, not changes in rents. The coefficients of rent appreciation are insignificant in all of his results. This provides evidence for potential price instability where movements in rent to price ratio are based on prices moving, not rents.

Chichernea et al. (2008) compare aggregate apartment capitalization rates between 34 US metropolitan areas and explain the geographical variation in aggregate capitalization rate of multiapartment properties, based on 2,116 transactions between 2000 and 2005. The controls for property characteristics include square feet, age, distance to the center of the city, location (metropolitan area) and time. The transaction data is clearly different from our study as it comprises observations where whole buildings of multiple rental apartments have been sold. Thus, the stock of housing is clearly different from our Finnish data, where owner-occupied and rented apartments can be located next door to each other. The same difference applies to Janssen et al. (2001) who had a similar dataset of complete rental properties.

However, the article by Chichernea et al. (2008) is relevant to this study, as it has a very rich set of metropolitan level variables to explain the variation in capitalization rates, which can be investigated using our data for individual apartments and houses as well. They model the expected growth rate on both demand and supply side factors. Demand side variables include employment growth, GDP growth, income growth and population growth. On the supply side, they test three separate indices used to proxy supply side constraints such as stringency of regulation and development process restrictiveness, which have been shown to positively impact housing prices. They also include measures for market liquidity and capital flow measures.

Expected effects are intuitively clear: higher future demand decreases CAP rate through higher expected appreciation rates, more supply constraints mean lower discount rates as rent and price growth is harder to match with more supply. Liquidity means less risk and it reduces capitalization rates as do capital flows from large markets towards smaller markets.

The results from Chichernea et al. (2008) suggest that the supply side constraints are powerful determinants for the variation in the apartment capitalization rates across metropolitan areas. However, the demand side factors receive little significance. Liquidity and capital flow measures seem to explain variation in the capitalization rate and their effect is negative as predicted.

Vacancy rate is clearly an important variable regarding the net rental yield, as all rent data should theoretically be adjusted downwards along with the vacancy rate, pushing net rental yields lower. There is potential variation in the vacancy rate across dwelling types and locations, but in the Finnish context the data availability is a major problem for studying the effect. In their analysis of user costs and rental equivalence, Garner & Verbrugge (2009) adjust their predicted rents with region-level vacancy rates and note the need for vacancy figures for more precise areas. Glaeser and Gyourko (2007) also note that there is potentially significant variance in the expected tenure length of different types of households, which would again result in variation in the net rental yields.

To our knowledge there is no credible research done to find out whether the capitalization rates vary between different residential building types or dwelling setups. In the Finnish context the legislative setting is homogenous between apartments (or condominiums in US terms) and multifamily properties built as row houses, two-family houses or even housing cooperatives consisting of multiple individual houses on a jointly-owned piece of land. In the case of single-family houses there is the unobserved and potentially significant variation in the lot size of the property that could have some sort of a role, but without sufficient data we are unable to model the lot size in the hedonic pricing functions.

2.3.3 Variables in hedonic value and rent functions

Malpezzi (2002) offers a thorough recap on the selection of variables for any hedonic model predicting rents or values of housing units. According to Malpezzi, the most important variables for a hedonic model are number of rooms, floor area, structure type, and age of the unit, socioeconomic neighborhood characteristics, distance to central business units or sub-centers of employment or education, time of data collection and possible tenant characteristics. However, Malpezzi notes that there are potentially hundreds of other housing characteristics that can be deployed in hedonic models. Sirmans and Benjamin (1991) review a vast number of studies regarding determination of market rent, covering all of the above variables. Correspondingly, Laakso (1997) goes through a number of studies on the hedonic pricing of housing units.

In the Finnish context, Laakso (1997) has investigated the demand for housing characteristics and urban housing prices, using transaction data of approximately 18,800 observations ending

1993. No related studies have been conducted regarding residential market rents in Finland. The following sections report briefly the effects of some variables on housing values.

Structure type and age

Laakso (1997) reports results on the structure type on dwelling prices in the Helsinki Metropolitan Area. He reports housing prices to increase monotonically with respect to dwelling size, and the elasticity is somewhat standard across sizes, except for size classes of around 200 square meters. It is worth noting that the prices have a modest jump around that range and this effect is signified in the case of multi-storey apartment buildings.

Laakso finds interesting results for the age variable. In Finland, the overall low general quality of planning, design and construction of the 1960s and 1970s suppresses prices the most compared to the referential decade of 1990. Looking back, house prices fell until the pre-war period and before the effect of the old vintage housing of the early century starts to increase prices from 1930s backwards. It is clearly evident, that age as a linear indicator of quality is of incorrect functional form when measuring housing values.

Most studies listed by Benjamin and Sirmans (1991) record similar relationship between rents and apartment age. Results on structure type are rare, as studies on rent mostly focus on apartment complexes in the US, and therefore, comparing the results to the Finnish context would be challenging.

Location variables

The basis for the modern microeconomic urban economics was created by studies of Alonso (1964) among others. The principal difference between urban economics and traditional microeconomics is in the role of location in the determination of demand, supply and price. Alonso presents a model of households' location choice which is based on traditional consumer theory. In his model, transport costs combined with distance to the city center are crucial factors in the determination of land rent and households' optimal location. After the arrival of hedonic pricing models (see e.g., Lancaster, 1966; Rosen, 1974), the structural properties of the dwellings and the characteristics connected with location and neighborhood variables are considered as components in the multidimensional basket of housing.

In the basic equilibrium regarding housing location, households maximize their utility between marginal transport costs and marginal savings from land rent costs, both of which are functions of distance to the central business district (CBD). The variation in households' bid rent functions, due to differences in incomes or preferences, explains why different kinds of households are located in different parts of the city, in other words, it explains why cities are segregated.

Four simplistic assumptions about cities are often made in the equilibrium studies (Fujita 1989). Firstly, cities are assumed monocentric with only one CBD, and the transport system is assumed dense in all directions and free of congestion. The city is also assumed round and uniform. Finally, the public goods and externalities are assumed not to exist. These assumptions are quite simple, and a vast number of studies have gone beyond the assumptions and extended the analysis to a very detailed degree, enhancing the basic model. However, as Laakso (1992) points out, even the enhanced spatial models are too restrictive and fail to explain housing prices and dynamics as a function of location only. Therefore, hedonic models incorporating the distance variables should be used.

Laakso (1997) uses several location factors to regress dwelling price. Most relevant factors to this study are the location to the central business district and location to the nearest railway or subway station. The results for distance to Helsinki center are unsurprising, although it is worth noting that for houses and row houses the distance effects on price are very large in the market for houses closest to the center. This is natural, because of the scarcity of houses within 10-15 minutes from the city center.

Vicinity of rail traffic is a more complex matter. Both accessibility and negative externalities are strongly connected to railway and subway stations. In HMA, the fastest public transport connections from suburbs to the city centre and several sub-centers are based on local railways and the subway. However, there are possible sources of externalities associated with railway stations. Many are located near road with heavy traffic, and in several cases, there are parking areas and feeder bus terminals close to the station. Hence, there are negative externalities caused by traffic in the form of noise, air pollution and accident risk. Unrest, crime and untidiness are also concentrated close to some stations.

Laakso's (1997) results generally signal that the positive accessibility factors outweigh the negative externalities for both the subway and railway networks. However, unless negative neighborhood effects are otherwise controlled for, the negative externalities for subway stations look extremely strong in a dummy variable model. Laakso concludes this to be a characteristic of neighborhoods in the immediate vicinity of the subway station. In general, subway has a further reaching effect to house values compared to the rail network.

Benjamin and Sirmans (1994) study the effect of mass transportation on apartment rents and find that in Washington D.C. rents fall 2.6% for every one-tenth mile decrease in the distance from the nearest subway station.

Another location variable, which is in the focus of this study, comes indirectly from the research on rental markets. As a natural clientele for rental apartments, students have been shown to impact rent levels. Benjamin and Sirmans (1991) report students' preference to live in close proximity to their campus having an impact on market rents. Ogur (1973) finds that rent is positively affected by college enrolment among other factors. Therefore, it may be worthwhile to incorporate a variable denoting distance from the nearest major campus in Finland.

Neighborhood variables

According to theoretical urban models concerning segregation, households prefer the homogeneity of the social structure of the neighborhood and want to live in an area where their own social group is well-presented (e.g. Fujita, 1989; Li and Brown, 1980). Laakso (1997) concludes that almost all hedonic studies find significant effects concerning the socio-economic structure of neighborhoods.

Laakso uses a variety of variables to analyze the effect of neighborhood-level socio-economic structure on dwelling values. These include education level, income level, share of owner occupied dwellings, unemployment rate, household size, social rental housing, population with foreign origin and crime rate. Because of high multicollinearity, Laakso proposes that only a small number of variables regarding social status should be used to avoid distorting the model specification. Laakso finds that the social status of the residential area has an extremely strong influence on housing prices in Helsinki.

Vacancy

Sirmans & Benjamin (1991) conclude that studies testing the effects of vacancy rates on rents have conflicting results. Nevertheless, most of the studies seem to find a negative effect on rent. In the related literature the researchers have also coined a concept of natural vacancy rate that changes between metropolitan areas. Rosen and Smith (1983) study vacancy rates and conclude that much of the variation in vacancy rates between cities results from differences in the natural vacancy rates rather than the degree of market tightness. High growth areas have higher natural vacancy rates due to additions from the supply side and lower rents.

2.4 Evaluation of housing market efficiency

As discussed above in section 2.2, the housing market efficiency can be analyzed using the basic model of housing values. Expectations about house price appreciation are important in analyzing housing market efficiency. Most empirical research testing the efficiency of real estate markets focuses on informational efficiency. In their classic paper, Case and Shiller (1989) show that markets adjust slowly to changing market conditions. They also show that, in the short-run, one can time the market and earn excess profits, which runs counter to the efficient market hypothesis. However, they conclude that it is nearly impossible to definitively prove whether the housing market is overvalued or not at certain point in time. Interestingly, they note that the historical time-series on implicit rents of owner-occupied houses is crucial to study these inefficiencies, because they see implicit rents as dividends for housing, which should be studied together with the price appreciation.

Inefficiencies with respect to observable demographics have been commonly in focus. For example, Mankiw and Weil (1989) argue that, in the US, there was an increase in demand for housing due to the Baby Boom in the 1970s, during which the inflation rate increased substantially. This increase was predictable well in advance. They state that the housing market probably should not be characterized as an efficient asset market in which prices reflect available information on future demand.

As some researchers have pointed out, predicting housing market bubbles has almost become a common past time in the US and around the world. Stiglitz (1990) provides a general definition of asset bubbles: “If the reason that the price is high today is only because investors believe that the selling price is high tomorrow – when ‘fundamental’ factors do not seem to justify such a price – then a bubble exists.” At least in the short run, the high price of the asset is merited, because it yields a return (capital gain plus dividend) equal to that on alternative assets. Case and Shiller (2003) define a housing market bubble as being driven by home buyers who are willing to pay inflated prices for houses today because they expect unrealistically high housing appreciation in the future. An analysis based on the NAC model can give interesting insights about bubbles. Even in its narrowest form where rents are compared with observed prices, it implies the appreciation rates expected by markets.

Oikarinen (2009) notes that because of their simplicity and appealing intuition to the general public, housing rent-to-price ratio and housing price-to-income ratio have repeatedly been employed, e.g., by credit institutions and the media to justify views concerning the sustainability of existing housing price level. However, as already discussed with the breakdown of the model, several of the variables are too vague and based on expectations to give accurate information on the state of the market as a whole. Long term changes in supply, demand, lot prices and institutional settings can have significant impact on the aggregate rent-to-price ratio, and the constant rent-to-price can shift up or down (diPasquale & Wheaton, 1996). Finally, as discussed earlier, due to differences in the quality of rental versus owner occupied housing, appreciation rates and rent-to-price ratios calculated from aggregated data have considerable measurement errors. Analyzing the issue of efficiency in the housing market using aggregated data may give biased results.

In the market for corporate equities, the forecasting power of the dividend yield (i.e. the ratio between dividend payments and the stock price) on future returns is a hypothesis that has a long tradition among practitioners and academics. The dividend-to-price ratio in the stock market is often used to examine market efficiency issues (see, e.g., Black and Scholes 1974; Fama and French 1988). In the housing market context, the rent-to-price ratio is treated as the real estate equivalent of a dividend-to-price ratio for equities. Even though calculation of total returns on the housing market is nearly impossible, it is insufficient to deem cross-sectional variation in net rental yield as mere variation in the future expected appreciation rates. Variations in periodical cash flows are too important to neglect when considering the total return available for housing investment.

2.5 Short description of the Finnish housing market

The Finnish housing market can be divided into two main sectors: the privately financed housing sector and the subsidized sector. After the rent control system was ultimately discontinued in 1995, the privately financed sector has been free to form prices and rents without restrictions (except for the part of increases to existing contracts). Rents and prices in the subsidized sector, controlled ultimately by governmental entities, will be left out of the scope of this study. Nevertheless, it is important to bear in mind the dynamics of how subsidized housing can influence the prices and rents of the privately financed sector, particularly as construction of subsidized housing generally decreases demand for privately owned rental units leading to decline in rents, housing values and private construction (DiPasquale and Wheaton, 1996).

According to Statistics Finland, over 60% of the Finnish dwellings are owner-occupied, while the rental market is internationally small at less than 40% of the total housing stock of 2.5 million dwellings. Approximately 1% of the homes are “right of occupancy” dwellings. Statistics Finland estimated the total market value of Finnish housing stock at roughly 200 billion Euros at the end of 2004.

According to Oikarinen (2007) the subsidized public sector and institutional investors own approximately half of the rental homes, and industry information provider Kiinteistötieto provides a figure of slightly less than 400,000 non-subsidized rental dwellings in Finland. Of these, some 50% are owned by small, mainly individual investors, while the rest are owned by institutions of varying levels of professionalism.

In the broader European perspective, the share of the Finnish rental market of the total housing stock is relatively small, which is likely due to a history of tax-incentivized home ownership. These particular tax questions have, from time to time, been in the focus of public discussion in Finland and are covered in section 2.2.4.

Historically, pension funds, insurance companies and not-for-profit foundations have formed the bulk of institutional investors in the housing market. However, significant players such as VVO and Sato, who have historically been involved in the subsidized sector, have increased their participation in the privately financed sector as well, and the market has also seen the formation of the first housing market focused property funds during the first decade of 2000s.

The funds investing in rental residential properties have taken the form of limited partnerships, while the Finnish legislative framework regarding the development of REITs is still lagging behind in the public tax discussion and has not been modified to resemble the common international models.

In the Helsinki Metropolitan Area, where the liquidity of the housing markets can be expected to be the greatest, the composition of the housing stock of roughly half a million dwellings differs somewhat from the national stock. As much as three quarters of the dwelling stock is located in multi-storey buildings and the share of privately financed apartments is approximately 40%. Of this, little less than 60% is owner-occupied (StatFin). According to Helsinki's official webpage[4], 39 % of all dwellings are privately financed, owner-occupied and 23 % privately financed, rental occupied.

Oikarinen (2007) provides a thorough historical perspective on the evolution of the Finnish housing market since the 1970s. Finnish housing markets have gone through major institutional changes during the past two decades. Firstly, in the late 1980s, the Finnish financial market was deregulated. Secondly, in 1993, there was a reform in the tax codes concerning the deductibility of mortgage interest payments in taxation. Thirdly, rent regulation was released in several stages during 1992-95. These institutional changes have had major impacts on the housing price dynamics. Furthermore, mainly due to increased migration from peripheral areas to the Helsinki Metropolitan Area (HMA) and to a couple of other centers in Finland, regional housing price development has diverged to a much greater extent since the early 1990s than earlier.

After the oil crisis of the 1970s, the housing prices increased nominally, but decreased on real terms until the start of 1980s. Following a few years of steady prices in the mid-80s, an immense rise in the housing prices began in 1987. This was to a great extent a consequence of the financial deregulation that took place in Finland, and as ceilings on average lending rates were abolished by the Bank of Finland, the availability of housing loans exploded in 1986 leading to a housing market boom.

---

[4] www.hel.fi

Figure 2:

Real price index for Finnish condominiums Q1/1970 – Q2/2009, 1970 = 100

Figure 2 illustrates housing price appreciation from 1970 to the second quarter of 2009. 1970 is the base year, set at the value of 100. The spike after year 1986 is often attributed to the deregulation of Finnish financial market, while the rapid decline in the early 1990s was simultaneous with a general period of recession in the economy.

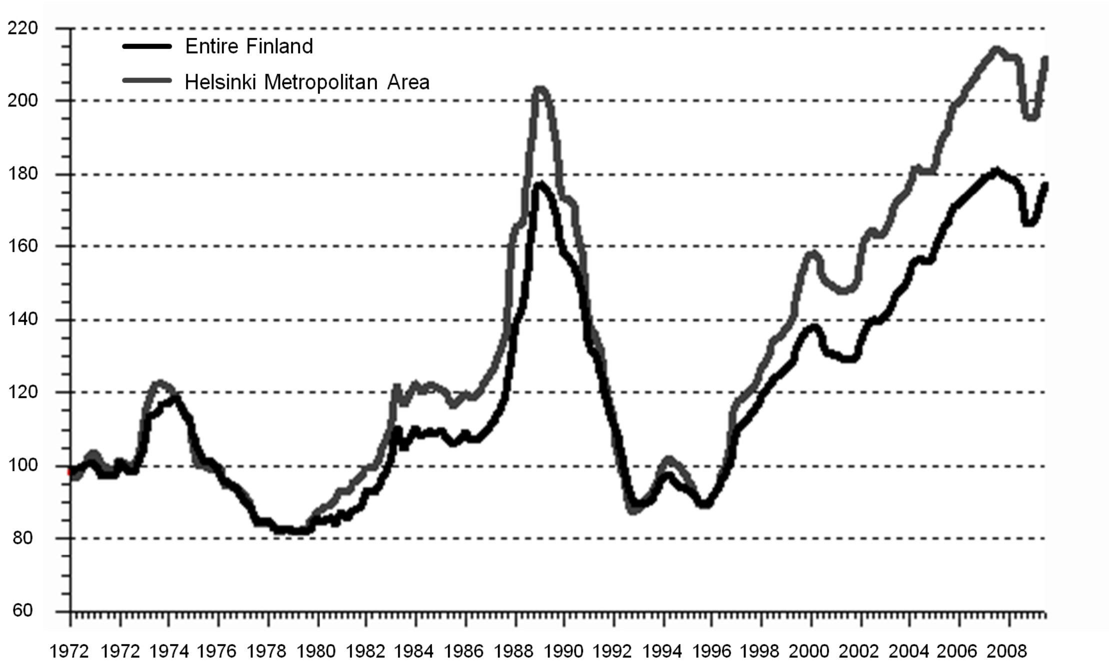


Figure 2 illustrates the housing price appreciation from 1970 to date. Following a real price increase of 58% from late 1986 to 1989 the market crashed by 50% before the end of 1992. The HMA was even more volatile as it fell over 66% and 57%, respectively. The most dramatic movements occurred in southern Helsinki, where the prices peaked a quarter later than in the rest of the country. After the recession of the 1990's in Finland, there has been a steep increase in prices, which has been led again by the HMA, which has experienced more rapid price increases than the rest of the country. The rise in the real housing prices during the 2000s has raised questions regarding the sustainability of the price development, and the ongoing global financial crisis and recession has again turned prices on a downward slope in the beginning of 2008. However, in the HMA, prices have this time proved more resilient than in the rest of the country.

2.5.1 Finnish household preferences regarding housing consumption

In this section we will go through the conclusions in Juntto's (2007) research on the preferences of Finnish consumers regarding their housing market consumption. The data is collected from 3,455 households in personal interviews and the overall representativeness is high. In the following, we will briefly summarize the themes of most interest to analysis of rental yields.

Overall, Finns have a strong preference (86% of respondents) for owner-occupancy as soon as they are financially in a position to take a mortgage. However, almost 40% of Finns live in rental apartments, and most of the low-income households live in rental dwellings. Owner-occupancy is amongst all survey participants considered a cheap alternative to renting, and a dwelling is also considered a good investment

Over a third of tenants in rental dwellings state that the primary reason for their tenure choice is the inability to buy a house due to low income or problems with saving the required mortgage down payment. This offers support for the argument that housing markets are in fact heavily segmented with low-income citizens flocking in the low-end of the housing market without an opportunity to freely choose between tenure choices. Also small niche-groups, such as students (94% of them tenants), are a recognizable part of the rental market.

The number of small households is increasing constantly as the average size of households declines. This macro-level trend should support the price appreciation and rental demand for the stock of housing with less rooms and smaller floor areas. On the other hand, for the wealthy deciles, the number of households living in detached houses has increased and detached houses are generally felt a more desired form of housing than before.

However, changes could occur rapidly, because Finns move quite often and over 25% of the survey respondents stated that they had planned moving in the near future (Juntto 2007). Housing decisions are made on comparing a rather small set of alternatives and decisions are mainly based on other than financial considerations.

3 Methodology & hypotheses

In this section we discuss the construction of the hedonic models employed in this study. We will first run through general issues on the choice of functional form and then proceed to choose a functional form to be used in this study.

3.1 Functional form

Much of research on hedonic models builds on the early work by Lancaster (1966), Rosen (1974) and Maclennan (1977). The models, as suggested by the name, are models which focus on analyzing the demand and prices for different sources of pleasure. These different sources of pleasure combine to characterize heterogeneous commodities in terms of their 'quality' (Sheppard 1999). In the analysis of housing, quality and pleasure should be understood, in a broad sense, as all of the attributes that make up the value of a house: location, size, condition, neighborhood, etc.

The standard approach to the analysis of housing prices by estimating a hedonic price function has been to adopt a parametrical approach. In this approach, a functional form for the price function is determined, after which a finite number of parameters are chosen and estimated to best determine the value of that function. These parameters reflect, or proxy, the hedonic attributes that form the demand for the house, and the coefficients represent estimates of the implicit prices of these attributes. The fundamental hedonic equation (Malpezzi 2003) is

$$R = f(S, N, L, C, T), \tag{9}$$

where R = rent (substitute V, value), S = structural characteristics, N = neighborhood characteristics, L = location within the market, C = contract conditions or characteristics and T = the time rent or value is observed. However, the actual functional form of the equation is subject to much debate. Economic theory does not place restrictions on the functional form of hedonic equations. However, the model must include land values, which depend on location (Sheppard 1997) and the equations should be convex (Jones 1988). They should also account for the number of public goods or amenities available (Parsons 1990).

In their 1981 paper, "The Choice of Functional Forms for Hedonic Equations", Halvorsen and Pollakowski were one of the first to propose the choice of functional form to be based on relevant statistical procedures, rejecting all of the most commonly used functional forms until then. They specify a highly general form that includes two different Box-Cox transformations and interaction terms. This quadratic Box-Cox functional form incorporates all other functional forms of interest as special cases and is thus highly flexible.

Cassel and Mendelsohn (1985) comment on the article by Halvorsen and Pollakowski. They provide an interesting discussion on the choice of functional form for hedonic price equations. They state that choosing the optimal functional form includes a tradeoff that depends on the purported use of the function. If predicting total house values is the main goal, the form that provides the best fit (typically the most flexible model) is quite appropriate, however this form may be suboptimal if information on the relative importance, or weight, of the different attributes is of interest. Though it may have less predictive power, some other functional form may be able to compensate for its shortcoming by providing more stable parameter estimates and a more intuitive interpretation.

Further commenting on the same subject, Cropper, Deck and McConnell (1988) investigate the issue of model selection. By simulation, they evaluate linear, log-linear, quadratic, linear Box-Cox and quadratic Box-Cox models on both predictive power and how well the model estimates marginal effects. Their analysis shows that in situations where some important variables were omitted or observed imprecisely, the simpler forms such as the linear Box-Cox, the linear, the semi-log and the double-log perform the best, while the quadratic and the Box-Cox quadratic forms performs the worst. As a conclusion, they advocate the Box-Cox linear form for having the smallest average bias, though the linear form produces the smallest maximum bias.

In his 2003 paper, Malpezzi reviews the major issues related to hedonic models and their applications. Regarding the functional form, he finds it "particularly problematic that theory yields little guidance to the functional form of the hedonic relationship." He compares the virtues of the linear form, the log-linear form, the translog form and the Box-Cox forms, but makes no conclusion about which is most suitable. Instead, he quotes Maclennan (1977) by stating that "the design of the pricing model should fit the purpose at hand."

In choosing the functional form to use in our paper, we will begin by performing explorative analysis using linear Box-Cox specification, allowing for different transformation parameters for the dependent variable and the independent variables (termed θ and λ respectively). The analysis will be done for a restricted sample (namely the Helsinki exact matched sample) in an attempt to ensure best possible comparability of fit. On the basis of this analysis, we will attempt to choose a functional form that provides a good combination of fit, stability and intuition. Additionally, we will use the linear form to verify the results and ensure that they provide intuitive explanations for the coefficients. For example Arimah (1996) and Laakso

(1992) arrive at using the double-log with similar reasoning (combination of interpretation, significance and stability). Laakso also performs Box-Cox tests and concludes that "the results did not change the main conclusions based on the results of log linear models."

3.2 Choice of functional form to be used in this study

In this section, we will choose a functional form by performing explorative analysis using the linear Box-Cox specification as discussed earlier in section 3.1.

The Box-Cox transformation is a parametric transformation method (Box and Cox 1964) that aims to ensure that the usual assumptions[5] of linear regression models hold:

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda\left(GM(y)\right)^{\lambda-1}}, if\ \lambda \neq 0 \\ GM(y)\log y, if\ \lambda = 0 \end{cases} \tag{10}$$

where λ is the transformation parameter and *GM(y)* is the geometric mean of *y*.

According to Draper and Cox (1969), even in cases where the power transformation does not give an exact normal distribution, the usual estimates of the transformation parameter λ will lead to a distribution that satisfies certain restrictions on the first four moments[6], meaning it will usually be symmetric. The transformation parameter can be optimized in two different ways: maximum likelihood estimation (MLE) and the Bayesian method. In both methods, λ, which minimizes the residual sum of squares, is obtained. There exist two special cases of the Box-Cox: When $\lambda = 1$, the transformation is reduced to a linear function, and when $\lambda = 0$, it forms a logarithmic transformation according to the L'Hôpital rule:

$$\lim_{\lambda \to 0} \frac{\left(x^{\lambda} - 1\right)^{\lambda}}{\lambda} = \lim_{\lambda \to 0} \frac{\frac{d\left(x^{\lambda} - 1\right)}{d\lambda}}{1} = \lim_{\lambda \to 0} x^{\lambda} \times \ln x = \ln x \tag{11}$$

---

[5] $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n)$

[6] (1) linearity of the relationship, (2) independence of the errors, (3) homoscedasticity (versus (i) time as well as the (ii) predictions) and (4) normality of the error distribution

We begin with the general form without interaction terms, proposed by Halvorsen and Pollakowski (1979), with the coefficients α_0 to α_i as well as the transformation parameters λ and θ to be estimated:

$$Y^{(\theta)} = \alpha_0 + \sum_{i=1}^{m} \alpha_i X^{\lambda} + \epsilon_Y \tag{12}$$

Definitions as follows:

Y = net rental yield

θ = transformation parameter for dependent variable

λ = transformation parameter for independent variables

X = the vector of independent variables

α_i = the vector of coefficients for the variables i

m = number of independent variables (12)

ε = error term

For this analysis, X includes twelve variables altogether: size of the apartment, distance from Helsinki city center, age (dummy), dwelling type (dummy) and annual dummies. All the chosen variables have broad support in literature. We use a restricted sample of exact matched cases in Helsinki (2140 cases), which is the highest quality subset in terms of amount of matches and data available.

As a result of the maximum likelihood estimation for the general form, we obtain values for λ = -0.235 and θ = 0.421. They are both significantly different from zero at the 0.001 level, with a log likelihood of 7159.9. We then perform the same estimation, but imposing a constraint where $\theta = \lambda$ and get λ = 0.041. These results are not significant at the 0.1 level, with a log likelihood of 7135.1. Both the likelihood ratio test and the pseudo R-squared test show that there is only a very small difference between goodness of fit for the two specifications in favor of the unconstrained specification. We choose the constrained form and choose to set λ = 0. Thus, as our study does not intend to forecast but rather explain, we favor the intuitive interpretations for the coefficients of the logarithmic (double-log) transformations (see equation (11)) to get equation (13), similarly to Laakso (1992) and Arimah (1996).

$$\ln Y = \alpha_0 + \alpha \ln X + \epsilon_Y, \quad (13)$$

where α is the vector of coefficients α_1 to α_i. If the constant and error term are defined as logarithms as well, the equation takes the double-log form:

$$\ln Y = \ln \alpha_0 + \alpha \ln X + \ln \epsilon_Y, \quad (14)$$

which is used is used as the regression specification in the regression analyses in section 5.4. By taking exponents of both sides, we obtain:

$$Y = \alpha_0 X^{\alpha} \epsilon_Y \quad (15)$$

The benefit of this form is that we can choose to either regress the net rental yield directly, or to regress its constituents separately, as follows:

$$R = \beta_0 X^{\beta} \epsilon_R \quad (16)$$

$$P = \gamma_0 X^{\gamma} \epsilon_P \quad (17)$$

where β and λ are the coefficient vectors for dependent variable rent (R) and price (P) respectively. When net rental yield is defined $Y = R/P$ (and thus (15) = (16) / (17)), we get:

$$Y = \ (\beta_0/\gamma_0) \times X^{\beta-\gamma} \times (\epsilon_R/\epsilon_P) \quad (18)$$

Where $\beta_0/\lambda_0 = x_0$, β_i - $\lambda_i = x_0$ and ε_R / $\varepsilon_P = \varepsilon_Y$. Separating the regression analyses for rent and price arrives at the same result but with the benefit of allowing for separation of rental yield into the constituents, giving possible further insight into the factors driving rental yield.

$$y_i = \propto_0 + \sum_{i=1}^{m} \alpha_i X + \epsilon \quad (19)$$

3.3 Hypotheses

Tian (2008), Garner and Verbrugge (2009) and Hargreaves (2005) have found that there exists significant variation in the residential capitalization rate between dwellings of different value. All of the authors have noted that the squared value of a house has a negative relation to observed rents. Therefore, our first hypothesis is:

H_1: There is a negative relation between the net rental yield and house value.

To test H_1, the data will be split into deciles according to dwelling price to see how net rental yields differ between the deciles. Additionally, the data will be plotted into a net rental yield on price scatterplot. Finally a simple regression test will be performed, with price and price squared as the independent variables and rent as the dependent variable.

If H_1 is true, it follows directly that the price and rent of a dwelling cannot be set according to the same formula. Moreover, for net rental yield to decrease for higher priced dwellings, one of the following must be true:

1. Potential buyers value some of the dwelling-specific characteristics[7] of higher priced dwellings more positively than renters do.
2. Potential tenants value some of the dwelling-specific characteristics of lower priced dwellings more positively than buyers do.

If on a general level either one of the above is true on an aggregate level, it implies that the elasticities in the hedonic price and rent functions differ. Therefore, we arrive at our second hypothesis:

H_2: Differing elasticities in the hedonic price and rent functions drive the variation in net rental yield.

H_2 will be tested using a regression analysis, where the dependent variable is either net rental yield, annual net rent or debt-free ask price and the independent variables are all the dwelling specific variables available in our dataset.

---

[7] The dwelling-specific characteristics can be either structural, location or neighborhood variables.

4 Data description

In this section we outline the data we use to analyze the variation in net rental yields. First, we summarize the data from Oikotie, the provider of the dwelling advertisements, and second, we go through the matching process; how the sample of matched pairs of rental and for-sale advertisement is derived. Then, we briefly describe the sources of other variables, present descriptive statistics for the full sample and subsets of data, and conclude the section by discussing the potential limitations of the data.

4.1 Our dataset

Our data is gathered from Oikotie, one of the two largest Finnish web-based marketplaces for apartments and properties for sale and for rent. Oikotie is owned by Sanoma Oyj, a Finnish media conglomerate. Our dataset consists of 31,864 matched pairs (used synonymously with cases) of advertisements for rental and ownership housing units during the time period 2002-2009.

The data has been compiled from the whole database of Oikotie advertisements, with 937,070 entries and 123 fields altogether (each entry is an advertisement and each field is a descriptive attribute to the entries, such as size of apartment, number of rooms, the date of the advertisement or a free text field). To our knowledge, the Oikotie database in its entirety has not previously been used for academic studies of housing prices, rents or rental yields.

The sample itself has been created by matching advertisements for rental dwellings with advertisements for ownership dwellings. There are two categories of matched cases; exact matches and similar matches. Next, we will describe the matching process and the characteristics of the two classes of cases.

4.1.1 Dwelling-specific fields in the Oikotie database

As mentioned above, the Oikotie database contains 123 fields altogether. However, not all of these fields are of interest in this study, and they were removed from the sample at an early phase. The fields used in this study are described in Table 1.

Table 1:

Oikotie data fields

This table presents the matching criteria that were used to match the rental advertisements to the for sale advertisements in the Oikotie dataset.

Description
Dwelling type (apartment / row-house or two-family house / single family house)
Zip code where the dwelling is located
The city where the dwelling is located
Street address
Year when the dwelling was originally built
Size of the dwelling (m^2)
Monthly rent of dwelling
The debt-free asking price of the dwelling
Maintenance and upkeep costs per month
Number of rooms
Description of dwelling
Advertisement publishing date
Advertisement publishing end

4.1.2 Matching process

The matching process was begun using the cleaned up version of the original database, which included 817,120 advertisements. Most of the removed cases were advertisements for subsidized housing, which are out of the scope of this study. Entries missing relevant fields (e.g. sales price or date of publishing) or showing indications of the data having been entered incorrectly (e.g. sales price over 100 million Euros) were removed from the data. The refined data includes 165 287 rental ads and 651 833 for sale ads. Figure 3 summarizes the matching process.

The matching was done according to some binding conditions with exact fit required as well as baskets of criteria and conditions where an approximate match was considered sufficient. First, a rental advertisement was taken, and then a for sale advertisement was chosen to match it. Table 2 summarizes the matching criteria and how the matching was done according to the given criteria.

Table 2:

Matching criteria

This table presents the matching criteria that were used to match the rental advertisements to the for sale advertisements in the Oikotie dataset.

Criterion	Match type	Description	Basket 1	Basket 2	Basket 3
Zip code	Exact				
Apartment type	Basket		Apartments	Row houses / two-family houses	Single-family houses
Floor number	Basket	Only for apartments	Bottom floor	Everything in between	Top floor
Number of rooms	Basket	Total number of rooms, (bedrooms + other)	One room	Two room	Three rooms or more
Construction year	Basket		Prior to 1961	1961-1983	After 1983
Floor area	Range	Maximum deviation 10 % (logarithmic)			
Advertisement date period	Exact (overlap)	At least one day of simultaneous time online			

After the initial matching was done, rental ads that matched the criteria for more than one for sale ads were analyzed further according to the refined matching criteria (Table 3) which also included searching for the closest match according to two criteria.

Table 3:

Refined matching criteria

This table presents the refined matching criteria that were used after the initial matching of the Oikotie advertisements if a single match was not identified.

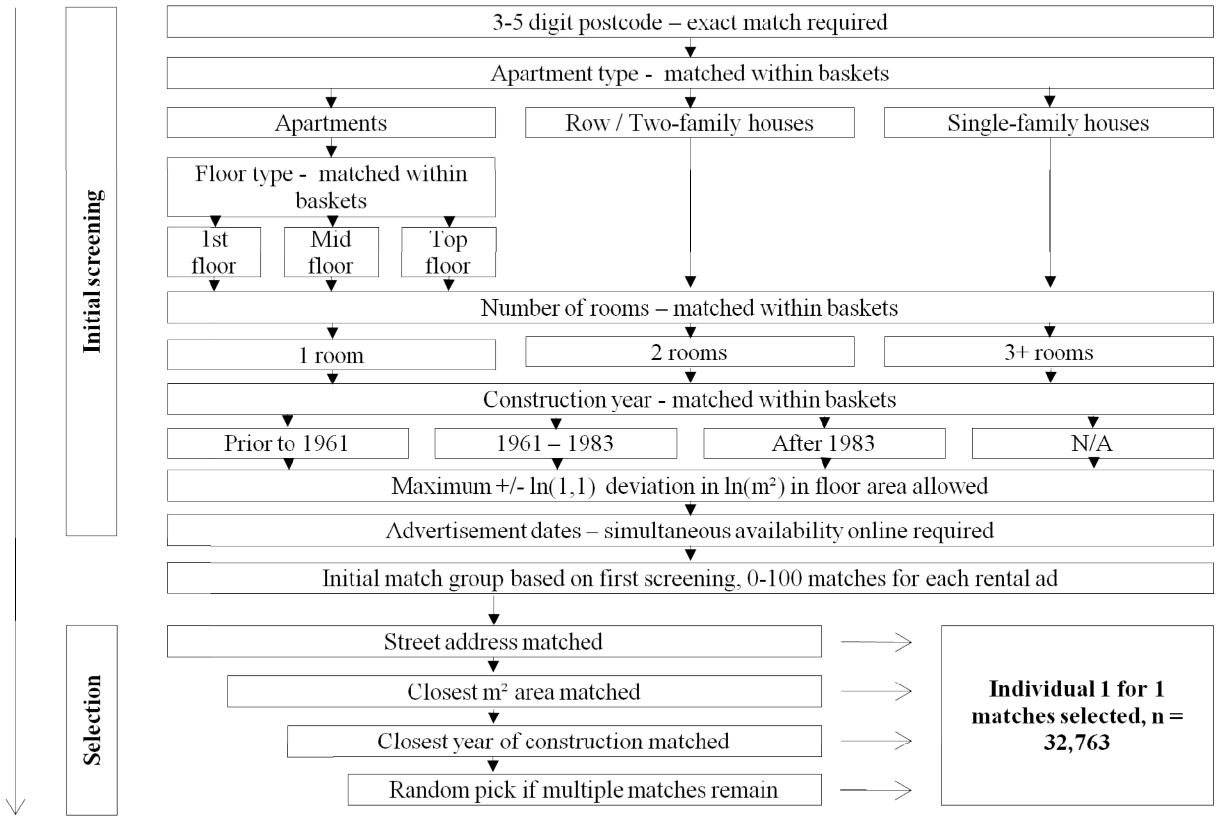
Criterion	Match type	Description
House address	Exact	If only one match found, that match was chosen, otherwise proceed to next step
Floor area	Closest	If only one match found, that match was chosen, otherwise proceed to next step
Year of construction	Closest	If only one match found, that match was chosen, otherwise proceed to next step
Random pick	None	One of the remaining matches chosen at random

After the two matching processes were completed, matches where two rental dwellings' advertisements had been matched to the same for sale advertisement were removed to make the matching procedure independent of which way it was done (rent to sale or sale to rent). The data included a sample of 32,763 matched cases. Of these, 4,747 were exact matches, where both all of the above mentioned criteria and the address (full address, not the house address which leaves ambiguity in multi-family dwellings) were exactly matched, suggesting that they were the same house that had been announced for rent and for sale simultaneously. In addition to the exact matches, 11,616 were matches that were on the same street.

Figure 3:

Matching methodology

Figure 3 summarizes the matching methodology used to create the sample of data. The initial screening, where all criteria are binding, is followed by the selection process, where the best possible unique fit is chosen.



4.1.3 Outlier analysis

As a final step in creating the sample, an outlier analysis was performed. Altogether 899 cases were removed. The outlier analysis consisted of three parts. First, cases that were the only entry within their zip code were removed (177 cases). Second, cases whose price or rent per square meter differed significantly from the rest in their zip code were removed (679 cases) and finally, a manual outlier analysis was done for the highest and lowest rental yields, removing cases (43 in total) where the free-text field revealed them to be unsuitable to the sample (e.g., flat let out on a per day basis, asking rent was per tenant, not flat).

4.1.4 Other dwelling specific data: Google Maps

In addition to the Oikotie data, dwelling specific data was retrieved concerning location variables. For this, the coordinates of all of the addresses of the apartments were obtained from Google's Maps service. The coordinates were then used to determine the variables presented in Table 4.

Table 4:

Google Maps data fields

This table presents the data fields obtained using Google's Maps service (http://maps.google.com). The distances between the locations are calculated using the haversine formula (see equation (20)).

Description
Distance to closest of the five metropolitan areas' city centers
Distance to the closest university campus (15 campuses altogether)
Distance to the closest commuter rail station (metro or train, only HMA)

The distances between the coordinates were calculated using the haversine formula, which measures the distance along the surface of a sphere.

$$haversin\left(\frac{d}{R}\right) = haversin(\varphi_1 - \varphi_2) + cos(\varphi_1)\cos(\varphi_2)\, haversin(\Delta\lambda) \quad (20)$$

where:

haversin = the haversine function

d = distance between the two points, along a great circle of the sphere

R = radius of the sphere

φ_1 = is the latitude of point 1

φ_2 = is the latitude of point 2

$\Delta\lambda$ = longitude separation

4.1.5 Area variables

In addition to the dwelling specific data from Oikotie and Google, two other sources of data are mainly used: Statistics Finland, which provides data that covers the entire Finland, and the City of Helsinki Urban Facts database, which includes more detailed neighborhood-level data on the Helsinki Metropolitan Area.

Statistics Finland[8], the Finnish public authority established for the purpose of gathering statistics, produces the vast majority of official Finnish statistics. We used Statfin for data on median incomes, town level unemployment and population growth.

The City of Helsinki Urban Facts[9] is a result of the merging of the Helsinki City Statistical Office and the City Archives in 1990. In this study we use City of Helsinki Urban Facts data on area average income, unemployment rates, share of dwelling types by tenure form (owner occupier, renter, subsidized renter or owner) and number of dwellings by tenure form. These data are only available for the Helsinki metropolitan area, about half of our total sample (15,478 cases).

4.2 Descriptive statistics

In this section, we will describe the dataset, which is formed as described in the previous section. We will show the distribution of the dataset as well as mean values of certain key variables. Finally, we will discuss some potential biases present in the data.

4.2.1 Annual distribution of data

As shown in Figure 5, the number of observations is heavily weighted towards the end of the sample period. This is mainly explained by the relative novelty of advertising dwellings on the Internet.

---

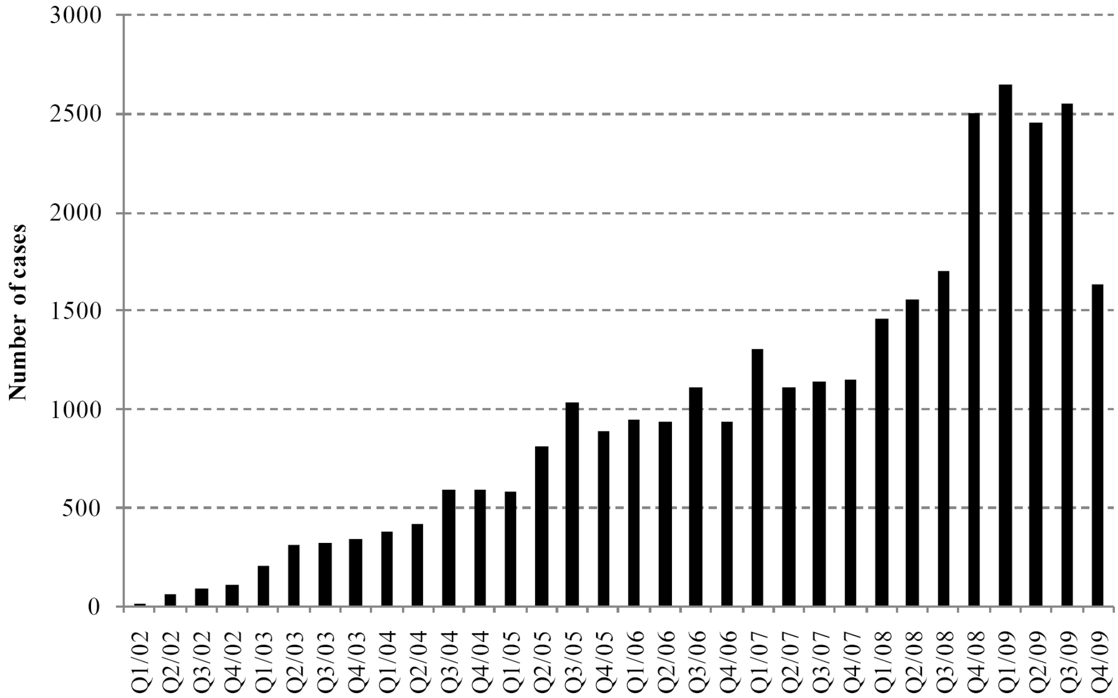
[8] http://stat.fi

[9] http://www.hel2.fi/tietokeskus/eng/index.html

Figure 4:

Quarterly distribution of the dataset

This figure shows the quarterly distribution of the cases within the period of analysis. As can be seen, data from the first years (2002-2004) is quite limited and most of the cases fall in the latter years. The data was retrieved halfway through Q4 of 2009, which explains the drop in observations.



4.2.2 Geographical distribution of data

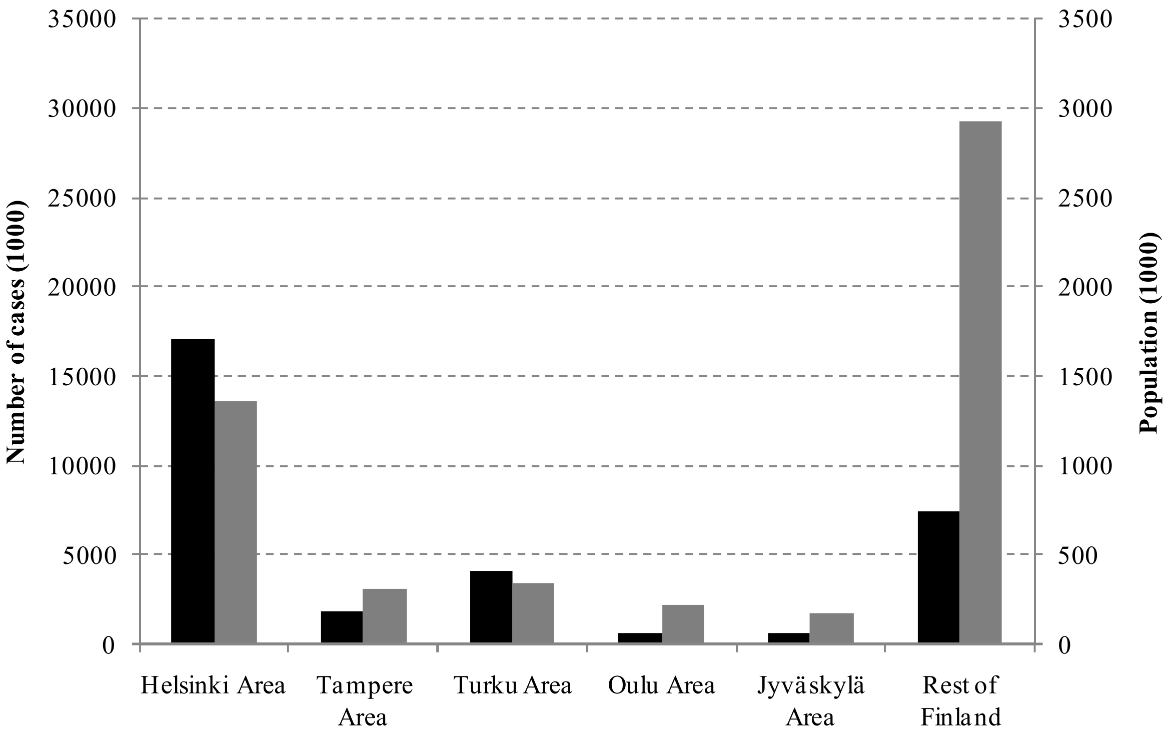
The data has been grouped into five metropolitan areas (Helsinki, Tampere, Turku, Oulu and Jyväskylä) reflecting the largest cities and metropolitan areas in the country[10] (Espoo is the second largest and Vantaa the fourth, but they are included in the Helsinki metropolitan area). Lahti is almost the same size as Jyväskylä, but was left out as an independent metropolitan area due to its proximity to Helsinki, as it would have been unclear which cities between Helsinki and Lahti fall into either category. As can be seen from Figure 5, the representation of the sample is best in the Helsinki and Turku metropolitan areas, while in the Tampere metropolitan area, the number of cases is slightly lower in relation to population, and in the Oulu and Jyväskylä metropolitan areas, the relation is significantly lower.

[10] Source: http://kunnat.net

Figure 5:

Geographical distribution of the dataset

This figure shows the geographical distribution of the cases within the period of analysis. The number of cases is on the left vertical axis and black bars, while population is on the right with gray bars. It can be seen that the Helsinki and Turku metropolitan areas are best represented in the sample, while the Tampere metropolitan area is slightly underrepresented and Oulu and Jyväskylä are strongly underrepresented.



4.3 Potential biases & limitations of data

In this section, we present the potential biases and limitations that are posed by using the advertisement dataset. The most obvious potential biases are caused by the use of ask prices instead of realized prices and the lack of data on vacancy rates. The major limitations are the lack of supply side variables and the inability to distinguish between institutional versus individual investors.

4.3.1 Ask prices

Because of using advertisement data, there are potential problems concerning the use of the asking prices and rents advertised online, and not using realized prices from, e.g., tax records or rental contracts. However, in the context of this study it is essentially irrelevant if a systematic bias exists among the individuals and institutions expressing their view on the value or rent of a certain dwelling. For example, if all prices were inflated by 10% to set prices higher before price negotiations, the rental yield curve would merely shift upwards, and not affect the variation between rental yields. Therefore, we assume that, even if the asking prices or rents were biased, the bias would be similar across all dwellings.

4.3.2 Vacancy rates

As discussed already in section 2.2.3, vacancy rates have been found to have significant but at times conflicting effects on market rents. Further precision about local vacancy rates would enable us to capture the realizeable net rental yield more effectively, and large deviations between vacancy rates of different dwelling types could potentially drive the differences in our observed net rental yields. Theoretically, correction for vacancy rate should of course be included in the net rental yield, but in our study, only cross-sectional variation of vacancy rates would have an effect on results, and therefore, we have not attempted to take them into account.

4.3.3 Supply side variables

As discussed earlier in section 2.3.2, Chichernea et al. (2008) suggest that the supply side constraints are powerful determinants for the variation in the apartment capitalization rates across metropolitan areas. However, this study does not incorporate supply side variables, such as data on construction permits, new housing construction starts, number of available lots for new construction, constraint indices or other variables, which could potentially capture cross-sectional differences in net rental yields. However, even finding such metropolitan-level data in Finland is difficult, and its application to the neighborhood-level would require even more detailed data.

4.3.4 Institutional versus individual investors

Sirmans and Benjamin (1991) find that professional property management has a significantly positive effect on rents. Our research makes no distinction whether the advertisements are entered by institutional investors, individual investors or households with any other idiosyncratic reasons for announcing a dwelling for rent or for sale. The origin of the advertisement could be hypothesized to have an effect on the ask prices, as property investors could be expected to pursue more analytical methods in their price setting decisions than ordinary small investors, who make a major share of the privately-financed rental housing markets. Secondly, if the price setting is done using this kind of systematic method, there should not exist significant cross-sectional variation of effective net rental yields. Support for the notion that institutional investors consider yields while setting prices and rents can be found in the data: by examining some of the newly built properties where multiple apartments are offered for sale and rent simultaneously, it can be seen that though the apartments are of very different price and rent, their net rental yield is exactly the same and has been set at a round figure.

5 Results

In this section we present the results of our analysis. We begin by showing the mean net rental yields over time, compared to interest rates and between housing submarkets in Finland. We then proceed to confirm the hypothesis of negative relationship between rents and dwelling prices. Finally, we present the results of the regression analyses and explain the sources of the variance in net rental yield extending beyond the mere rent-value relationship.

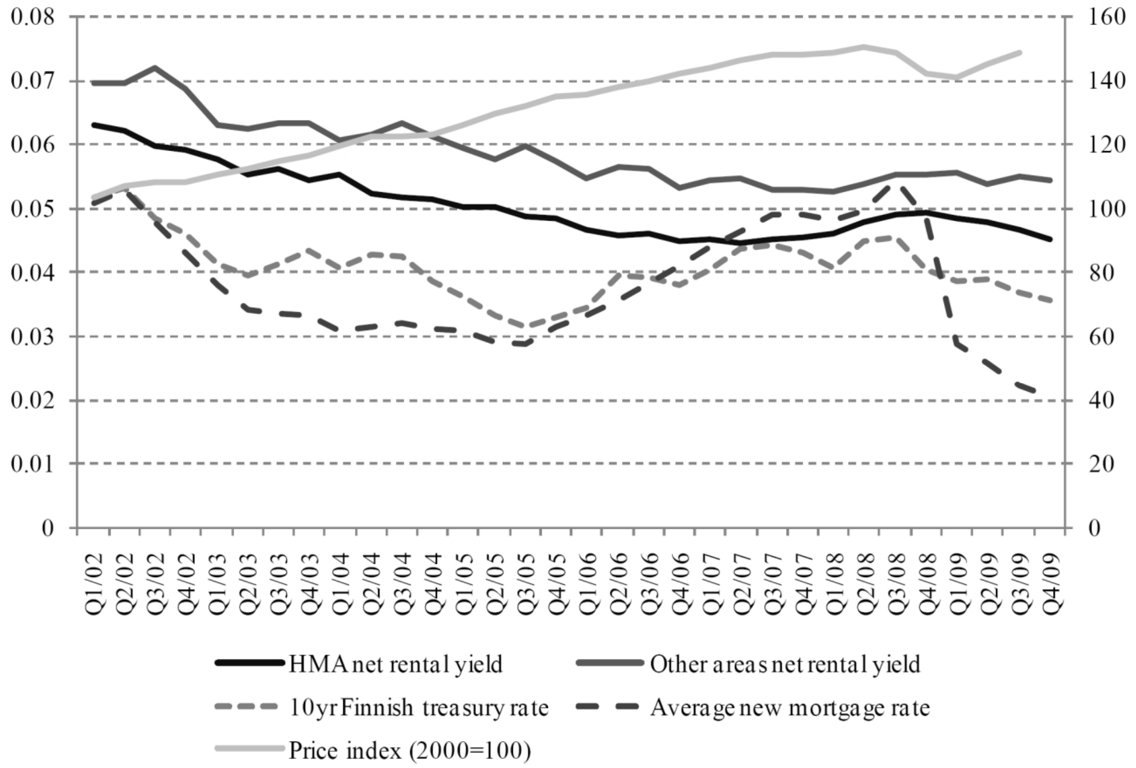
5.1 Mean net rental yield over time

Figure 6 shows that mean net rental yields have generally decreased in Finland over the period stretching from 2002 to 2009. As housing prices have continuously increased during the observation period, one can interpret the overall decline as the result of rents falling behind the high price appreciation markets. Until the second half of 2005, the net rental yield followed the decline in the long-term interest rate somewhat closely. The recent sharp decline of short term interest rates and simultaneous increase in house prices has pushed the net rental yield lower again.

Figure 6:

Quarterly mean net rental yield

Figure 6 shows the quarterly mean net rental yield for Helsinki and other areas from 2002 to the third quarter of 2009. Other series on the left hand side axis include the average new mortgage rate (mainly based on 3 and 12 month euribor rate and a mortgage default premium) and the 10yr Treasury rate. Price index of old apartments and row houses is plotted on the right hand side axis.

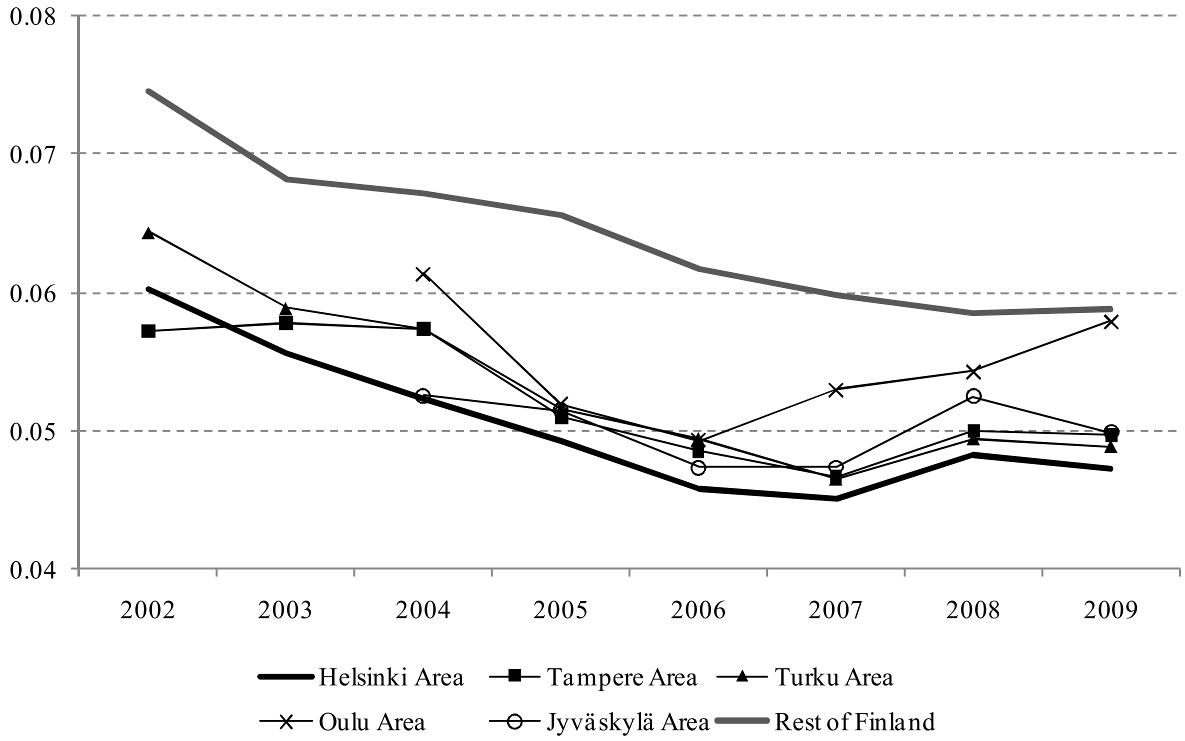


From Figure 6, one can note that during the 2002-2008 period there has been relatively long periods when the out-of-pocket user costs have been well below the gross rents. This can be interpreted to result from the spread between the new mortgage rate and net rental yield. During the four years before 2007, it was generally much cheaper to own than rent. In the Helsinki region, the mortgage rate crossed the net rental yield at the beginning of 2007, implying owner occupancy to be more costly tenure for households buying into the market at that time. However, Himmelberg et al. (2005), among many others, argue that it is more fruitful to compare net capitalization rate with the long term interest rate, as done in Figure 6. Bearing in mind the average mortgage premium of 50-150 basis points, the deviation of the synthetic long-term mortgage rate from the rental equivalence is not at all dramatic.

Figure 7:

Mean net rental yield by metropolitan area

This figure shows the annual mean net rental yield from 2002 to the third quarter of 2009 by six different metropolitan areas. Due to low number of observations, data for Oulu and Jyväskylä is not shown prior to 2004.



Following Figure 7, there is clear graphical evidence that the size of the market decreases capitalization rates. The Helsinki metropolitan area offers consistently lower mean yields, and areas outside the 5 main metropolitan areas have yields almost 1.5% higher than those in Helsinki. The recent increase in capitalization rates in Oulu seems somewhat surprising, but due to a relatively small number of observations (prior to 2008) this is hard to analyze. There are only around 50 recorded cases in the Oulu region during 2005 and 2006, which makes potential data error plausible.

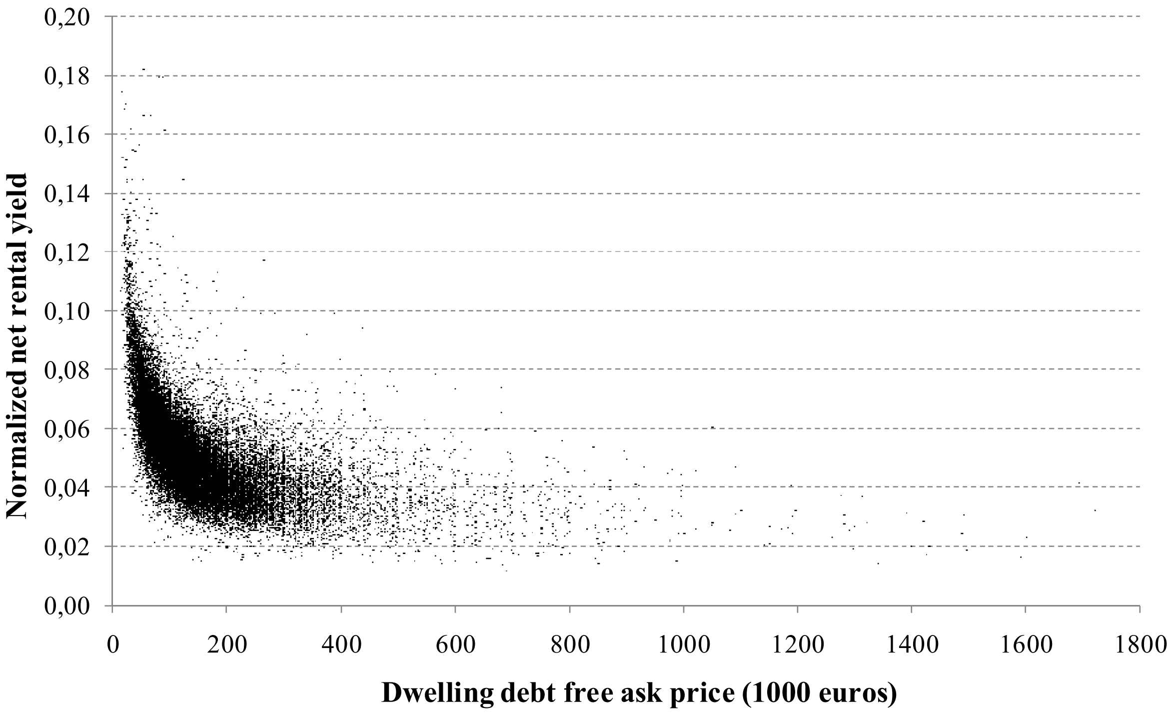
5.2 Net rental yield as a function of dwelling price

As discussed in section 2.3.1, net rental yield and asset value have been found to have a negative relationship. Figure 8 plots the normalized net rental yield against the debt free price in thousands of euros for all of the matches in our sample. Graphical support for the convex relationship is seen in the figure. Net rental yields tilt heavily towards the low-end of the market with a long tail of low yields extending to the high-value end. The basic form of the relationship looks logarithmic at a glance. In addition to the general shape of the curve, it is striking to note the complete absence of high-yielding assets in the very expensive end of the market. Low-rent bargains are available throughout the value axis, but even the lowest yields rise when approaching the less expensive end.

Figure 8:

Net rental yield as a function of dwelling price

Figure 8 plots the net rental yield against the debt free ask price in thousands of EUR. Net rental yield observations from 2002-2008 are normalized according to quarterly average yields. Altogether 31,864 observations are included.

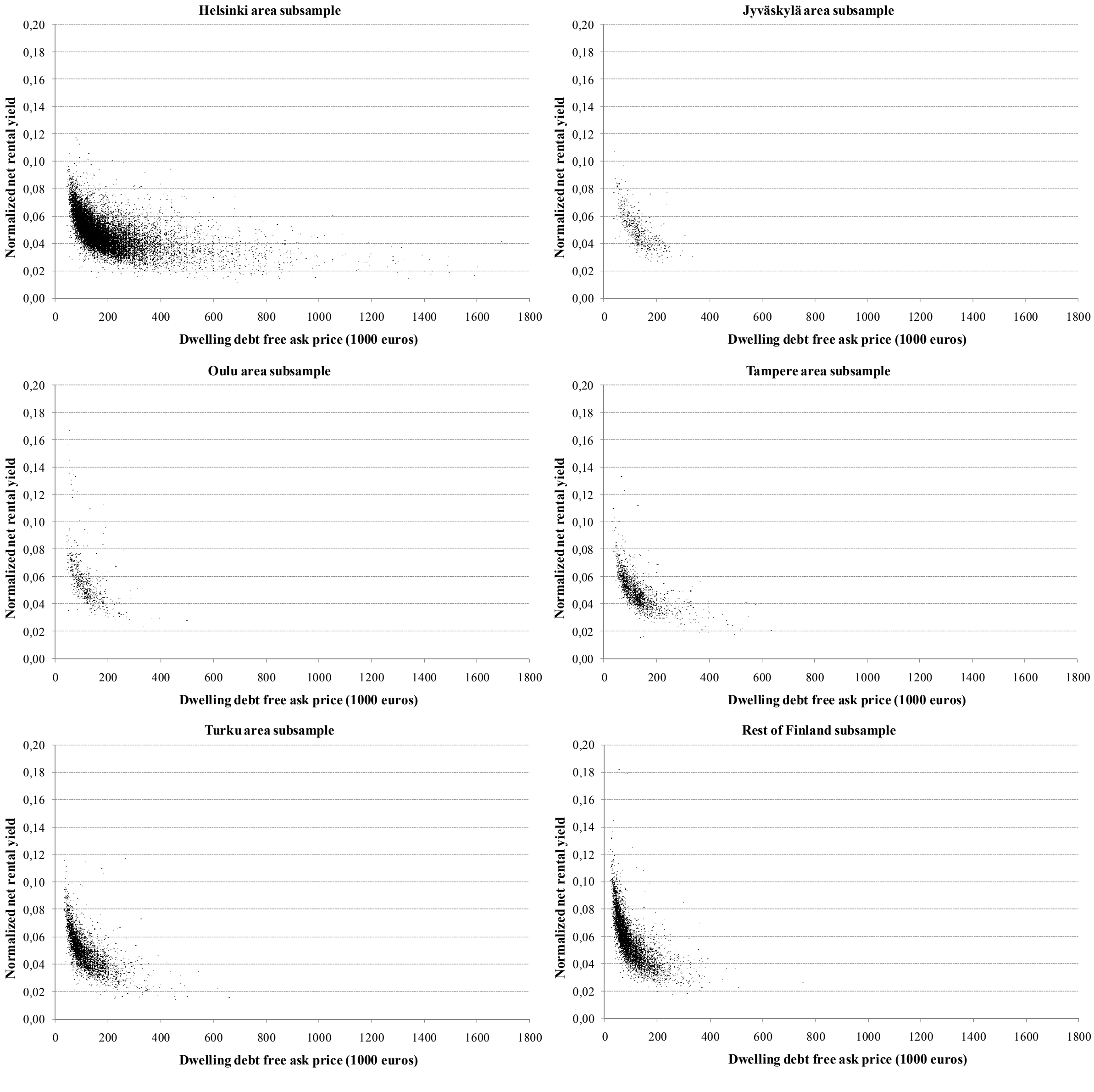


Furthermore, as can be seen from Figure 9, the phenomenon is consistent when the dataset is broken down into the six different geographical areas. It should be noted that in areas outside of Helsinki, there are less expensive dwellings due to land values being lower, but nevertheless, the yields fall strongly as debt free ask price increases. This gives further support for the notion that the variation is not due to differences between geographical areas but is inherent in the Finnish housing market generally.

Figure 9:

Net rental yield as a function of dwelling price in different geographies

The six panels in Figure 9 plot the net rental yield against the debt free ask price in thousands of EUR for the six geographical subsamples Helsinki, Jyväskylä, Oulu, Tampere and Turku areas as well as Rest of Finland. Net rental yield observations from 2002-2008 are normalized according to quarterly average yields. Altogether 31,864 observations are included, with 16,980 observations in Helsinki area, 594 in Jyväskylä, 631 in Oulu, 1,705 in Tampere, 3,575 in Turku and 5,134 in Rest of Finland.

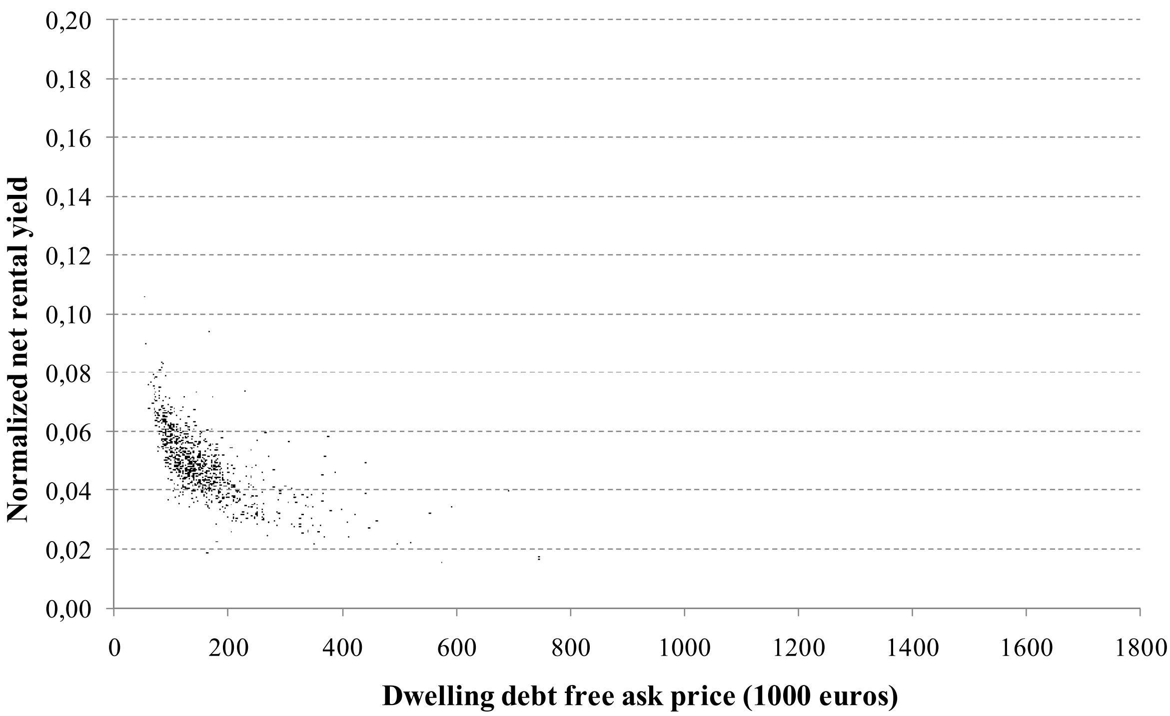


Breaking the sample down even further, the phenomenon shows consistency even within very small geographical subareas. Figure 10 shows the distribution of net rental yields as a function of debt free ask price in Kallio, an urban district of Helsinki located near the city center. Kallio (00530 Helsinki) was chosen because it is geographically a small and homogenous area, where the land prices are not likely to vary greatly and because the sample contains a large amount of cases from that area.

Figure 10:

Net rental yield as a function of dwelling price in Kallio, a Helsinki district

Figure 10 plots the net rental yield against the debt free ask price in thousands of EUR for Kallio, an urban district in Helsinki. Net rental yield observations from 2002-2008 are normalized according to quarterly average yields. Altogether 933 observations are included.



The complete dataset is divided into value deciles in Table 6. Examination of the mean yields lends further support to the hypothesis that the yields decrease as dwelling prices increase. Performing ANOVA analysis for differences in all means yields an F-statistic of almost 4000, hence the means are statistically significantly different from each other. The decline in mean yields is monotonic and consistent, although the minimums and maximums do not show a similarly consistent pattern. However, considering our sample size, the latter observation is not surprising. Importantly, standard deviations do not vary much either.

Average net rental yield for the least expensive decile is 4.0 % higher than for the most expensive decile, and a shift from dwellings costing 100,000 euros (4th decile) to over 200,000 euros (9th decile) decreases the net yield by 1.3%. In (pre-tax) money terms this translates to monthly savings of over 200 euros for the tenants of expensive dwellings, or similarly, smaller implicit cashflows earned by the owner-occupiers.

Table 5:

Mean net rental yield by dwelling price deciles

Table 5 below presents descriptive statistics of mean normalized net rental yield by dwelling price deciles.

	Debt free price (000 EUR)				Normalized Net Rental Yield			
	Min	Max	Mean	StDev	Min	Max	Mean	StDev
1st decile	14	59	45	10	0.031	0.188	0.080	0.017
2nd decile	59	76	68	5	0.023	0.181	0.065	0.011
3rd decile	76	90	83	4	0.022	0.199	0.060	0.010
4th decile	90	107	98	5	0.021	0.167	0.056	0.009
5th decile	107	123	115	5	0.026	0.158	0.053	0.009
6th decile	123	140	132	5	0.016	0.122	0.050	0.009
7th decile	140	165	152	7	0.016	0.117	0.048	0.008
8th decile	165	198	180	10	0.019	0.121	0.046	0.009
9th decile	198	269	229	20	0.017	0.121	0.043	0.010
10th decile	269	1720	409	170	0.013	0.109	0.040	0.011

To further support the argument for the negative relationship between net rental yield and value, we run the following OLS regression for net rent on the debt free price of the dwelling:

$$R_i = \propto + \beta_1 P_i + \beta_2 P_i^2 + \varepsilon \tag{21}$$

where R_i is the annual net rent of the dwelling, P_i is the debt free price of the matched dwelling and ε is an identically and independently distributed disturbance. Previous literature discussed in 2.3.1 has found β_1 to be positive and β_2 to be negative.

Table 6 presents the results for the regression analysis. Squared prices receive a negative coefficient as expected, and all results are significant at the 1% level. In addition, one can observe a decrease in the baseline rent as the number of rooms increases: the decreasing constant term in columns 1-3 suggests that the "base" rent falls slightly from 160 to 148 euros per month as the number of rooms grows from one to three or more. The squared price coefficient is smaller in the complete sample, which is likely to be attributable to more distributed observations and shorter relative range. The results for the exact match sample are very similar to those for the whole sample, suggesting there are no significant biases in the whole sample attributable to the slightly different matching methodology.

Table 6:

Regression results: annual net rent on price and price^2

Following equation (21), Table 6 presents the regression results of regressing annual net rent on dwelling price (1000 EUR) and price (1000 EUR) squared. Columns 1-3 represent separate regressions for apartments with one room, two rooms and three or more rooms, and the fourth column the regression for the whole data. The fifth column represents the regression for the data set of exact matches only. The figures in parentheses represent t-values. All coefficients are significant at the 1 % level.

	1 room	2 room	3+ rooms	Combined	Exact matches
Constant	1909.04	1876.27	1779.96	1810.20	1990.90
	(66.49)	(57.46)	(25.08)	(78.21)	(45.51)
Price	35.57	34.78	36.96	34.94	33.36
	(90.39)	(97.35)	(88.40)	(190.47)	(67.21)
Price^2	-0.0193	-0.0171	-0.0076	-0.0058	-0.0086
	(-17.39)	(-21.88)	(-17.74)	(-25.52)	(-8.80)
n	31864	31864	31864	31864	4747
R^2	0.72	0.73	0.75	0.79	0.79

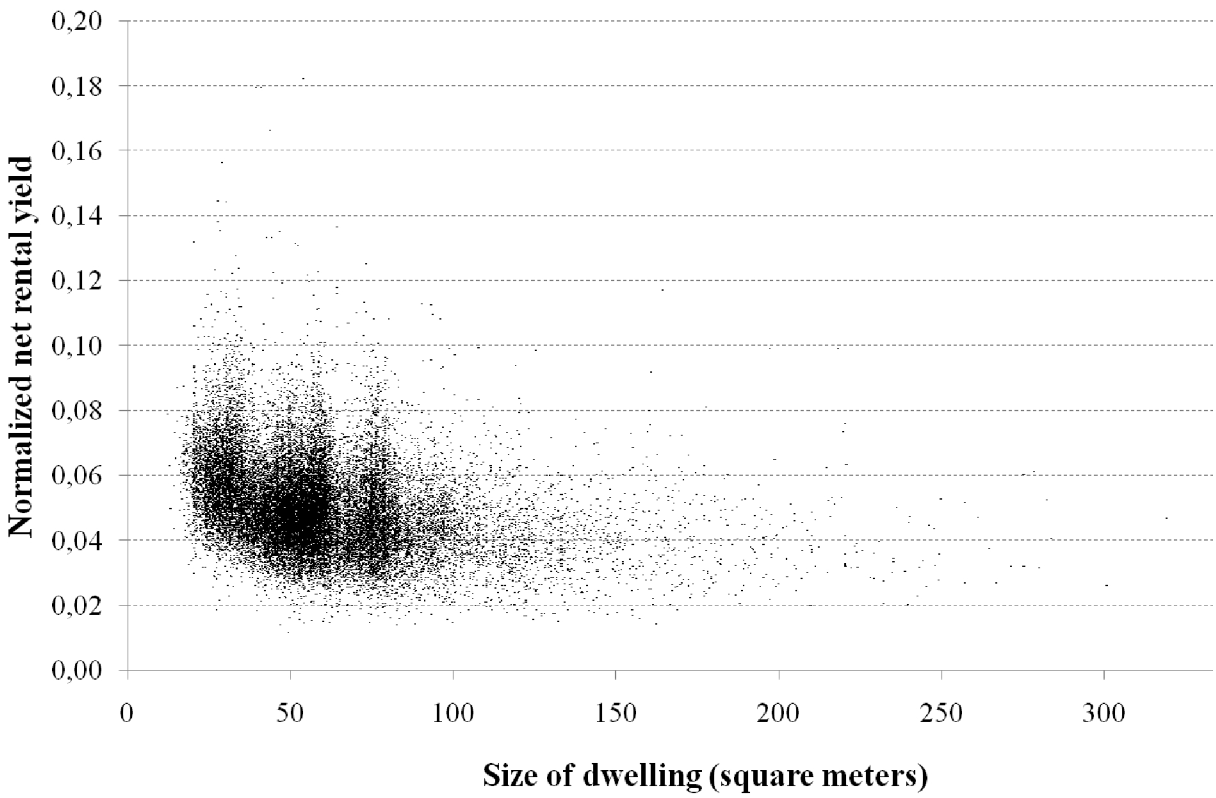
5.3 Net rental yield as a function of dwelling size

Because dwelling price and dwelling size can be expected to show a strong positive relationship, it is natural to assume that net rental yield also decreases when dwelling size increases. Figure 11 depicts the distribution of net rental yields as a function of dwelling size in the whole of Finland.

Figure 11:

Net rental yield as a function of dwelling size

Figure 11 plots the net rental yield against the dwelling's size in square meters. Net rental yield observations from 2002-2008 are normalized according to quarterly average yields. Altogether 31,864 observations are included.



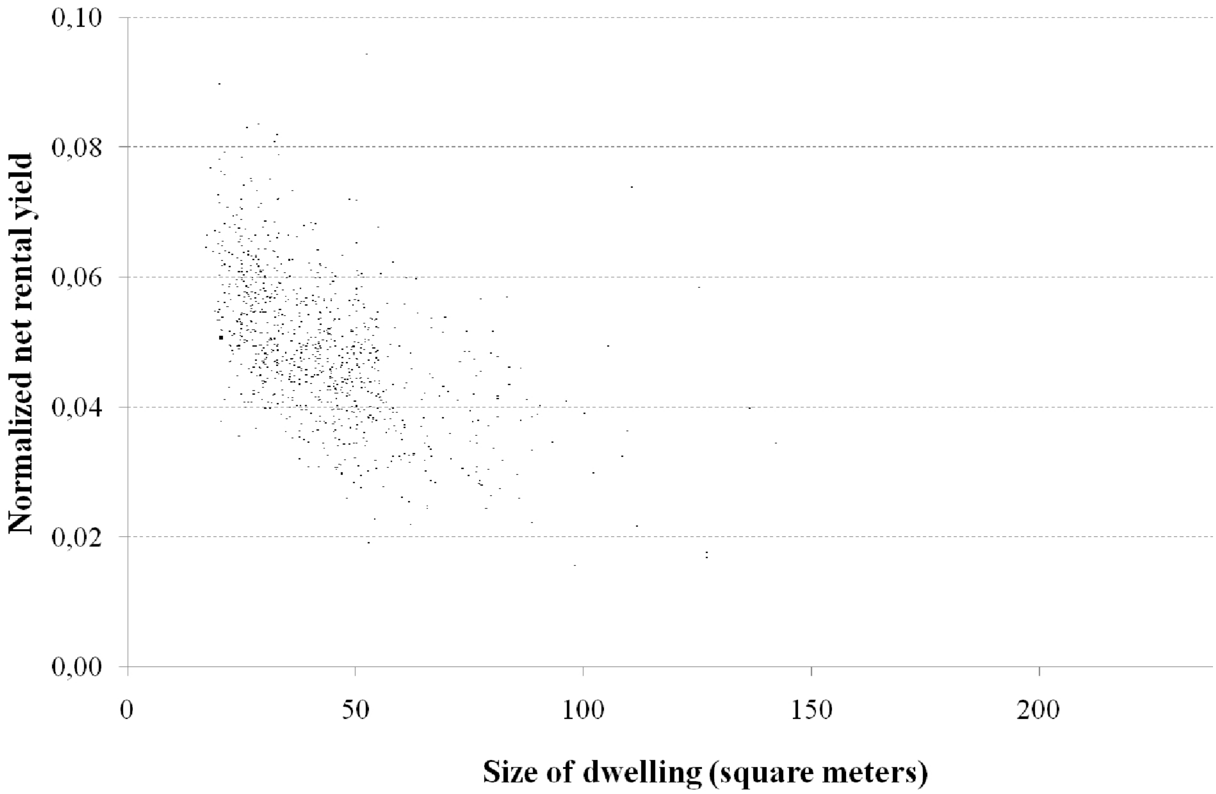
It can be seen that dwelling size does have a negative relationship with net rental yields, but that it is not as clearly decreasing as the relationship with debt free asking price. Due to the fact that land prices vary so greatly across the country and even single cities, the price of dwellings of the same size can vary greatly. This would suggest that dwelling price is a stronger driver of rental yields than size.

On the other hand, within a single zip code, land prices can be expected to be relatively stable. Therefore we again turn our attention to Kallio in Helsinki. Figure 12 depicts net rental yield as a function of dwelling price in Kallio and as can be seen, net rental yields fall as dwelling size increases. As the value of the lot that the (in Kallio, predominantly apartment building) structures are on is divided proportional to the size of the dwelling, a larger dwelling implicitly contains a proportionally equal amount of land value as a smaller one. Therefore, it seems quite unlikely that land leverage, or the fact that land does not depreciate, would explain more than a very modest part of the variation in rental yields.

Figure 12:

Net rental yield as a function of dwelling size in Kallio, a Helsinki district

Figure 12 plots the net rental yield against the dwelling's size in square meters for Kallio, an urban district in Helsinki. Net rental yield observations from 2002-2008 are normalized according to quarterly average yields. Altogether 933 observations are included.



5.4 Determinants of rents, prices and net rental yield

Although rents and prices both can be modeled by hedonic models, as discussed earlier, the coefficients for the models need not be exactly the same for both dependent variables. This also results in variation of the net rental yield, the quotient of rent and price. Tables in the next section present the results of regression analyses where the dependent variable is one of these three factors. As mentioned earlier, all regression specifications follow equation (14).

This section will present some final information on the regression variables and interpreting them.

In the regresssion analysis, the sample size has been reduced to 27,719 for the whole country dataset and 15,129 for the HMA sample due to unavailability of data from either Statistics Finland, Helsinki Urban Facts or on location (Google Maps).

5.4.1 Variables in common to both samples

All of the dwelling-specific variables in the regression analyses shown in this section are the same for both samples. Continuous variables that are in common to both are dwelling size as well as the distance variables.

The dummy variables include a count of observations in each dummy class. The reference classes for the dummy variables are as follows:

- Built before 1963 in the age dummy class
- Floor in between (not top or bottom floor) in the floor number dummy class (not apartment is also a part of this class, meaning the dwelling is either a row house, two-family house or a single family house)
- Lot is owned in the lot ownership status class
- Building has no lift in the lift dummy class

5.4.2 Variables specific to each sample

In addition to the metropolitan area dummies, there is one continuous variable specific to the whole country sample:

- Town level unemployment rate (unemployed / workforce)

In the Helsinki Area sample, the three sample-specific continuous variables are:

- Area average income for areas roughly equivalent to zip codes
- Area median advertisement period, the median duration of the publishing period of a dwelling advertisement on Oikotie in the area (as a proxy for liquidity / vacancy)

- Turnover of the dwelling type in the area: the number of rental advertisement in 2008 to total number of rental housing of the same type in the area (as proxy for liquidity / vacancy)

5.4.3 Interpreting the regression results

As demonstrated by equation (14), the regression results are OLS regression coefficients for the logarithms of each of the variables, with dummy variables entering into the regression either as 1 or e^1 (~2.718). An intuitive interpretation of the results can be gained by taking exponents of both sides of equation (14), arriving at equation (15), where the regression coefficients are the elasticities of each of the variables. A negative coefficient means that an increase in the parameter value will decrease the rent, price or net rental yield and vice versa.

As discussed in section 3.2 and equation (18) in particular, an attractive property of the double-log specification is the translation of the coefficients for net rent and price to net rental yield. The coefficient for net rental yield is the coefficient for net rent minus the coefficient for price (with the exception of the constant, which is the quotient). However, the coefficients of net rental yield have been analyzed in separate regressions to ensure accuracy.

5.5 Regression analyses

In this section we will first interpret the results of the whole country sample and then compare them to the results of the HMA sample. Finally we will evaluate the differences more qualitatively.

5.5.1 Net rental yield in the whole country sample

Table 7 depicts the regression results for two OLS regressions with double-log specifications, with net rental yield as the dependent variable and the whole country sample as observations. In the table, nearly all coefficients are significant at the 1 % level (and even at the 0.1 % level), so no asterisks to indicate significance have been added. The first column of coefficients shows the first regression with no area variables, while the second column includes an the town level unemployment variable.

Table 7:

Whole country sample: net rental yield

Table 7 shows regression results for the whole country sample using a double-log regression specification, following equation (14). The dependent variable is net rental yield, and the coefficients in the first column include dependent variables without area specific continuous variables. The rows with a count (n of dummy) are dummy variables. The rightmost column includes town level unemployment, a zip code specific variable. As all coefficients except for the top floor dummy are significant at the 1 % level, asterisks marking significance have been left out. The effects of advertisement date (all significant at 1 % level except for 2008, which is expectedly very close to the reference class 2009) and metropolitan area (all significant at 1 % level) are displayed in Figure 6 and Figure 7 respectively and therefore marked controlled here.

Variable	n of dummy	Coefficients (t-stat in parentheses below the coefficient)	
dwelling size		-0.265	-0.264
		(-78.45)	(-78.11)
distance from center		0.033	0.034
		(24.23)	(24.54)
distance from university		0.034	0.034
		(23.01)	(23.24)
built 1962-1983	10605	0.064	0.067
		(17.86)	(18.50)
built 1983-2009	7614	-0.105	-0.101
		(-25.69)	(-24.75)
not apartment	2828	0.115	0.119
		(22.54)	(23.19)
bottom floor	3829	0.019	0.019
		(4.86)	(4.92)
top floor	4152	0.008	0.008
		(2.19)	(2.22)
lot is rented	571	0.018	0.017
		(4.27)	(4.23)
building has lift	3422	0.013	0.014
		(4.47)	(4.51)
town level unemployment			0.051
			(6.15)
advertisement date		Controlled	Controlled
metropolitan area		Controlled	Controlled
constant		-2.141	-2.004
		(-157.61)	(-76.68)
Adjusted R-squared		0.375	0.376
n		27719	27719

The regression analysis done for the whole country sample of 27,719 observations gives strong support for the hypothesis that rental yield is not constant throughout the housing market, but that both dwelling specific as well as area characteristics have an effect on it. The full model's explanatory power, measured by R-squared, is 0.376, compared to 0 were the null hypothesis to hold. The coefficients are analyzed in detail in the following list.

i) *Dwelling size*

Dwelling size, which is naturally highly correlated with the value of the dwelling, has the largest absolute value and has a strongly significant negative coefficient.

ii) *Distance from center / distance from university*

Distance from the city center and from the closest university both have significantly positive coefficients, meaning that the closer the dwelling is to any of these, the lower the rental yield. This is also in line with the hypothesis of net rental yield decreasing with value, as closeness to such locations typically raises the value of a dwelling.

iii) *Building age*

The two dummy classes representing the age of the building are also in line with the hypothesis that increased value of a dwelling decreases net rental yield: for the years 1962-1983, the era of lower quality building, the coefficient is significantly positive, while for 1983 onwards (new dwellings) the coefficient is significantly negative.

iv) *Row or single-family house (not apartment) / dwelling floor (top/in between/bottom)*

Interestingly, the dummy for not apartment (meaning the dwelling is either a row house, a two-family house or a single-family house) has a significantly positive coefficient. This finding is confirmed by a difference of means test (not presented in this paper) within some of the most observation-rich zip codes: apartment dwellings of the same size and even value seem to offer a lower net rental yield than respective row-houses and single-family houses.

As may be expected, the bottom floor dummy has a significantly positive coefficient. Though its economic significance is limited, it is noteworthy that the effect on price and rent of being in the bottom floor is indeed significantly negative (Table 12 and Table 13). The top-floor dummy's coefficient is also positive, which is slightly unexpected, but its economic significance is lower still than that of the bottom floor dummy.

A look at Table 12 and Table 13 shows that being on the top floor actually has a negative effect on both rent and price, which would be very surprising. As for the HMA sample this is not the case, it seems that this may not be a real phenomenon but rather a bias in the data, perhaps caused by the lack of a combined dummy for both having a lift and being on the top floor. However the limited amount of observations with both of these conditions prevents us from performing this kind of an analysis with this dataset.

v) *Rented or owned lot*

The dummy for rented lot is significantly positive, which is in line with the theory of land leverage explaining higher yields (assuming that rented lots are priced too low in terms of their real value), but its economic significance is limited, at around the same level as the bottom floor dummy. However, due to the low availability of data on the lot being rented or owned, these results are not as reliable as the other dummy classes. (Table 12 and Table 13 show that the signs for rent and price are also in line with the land leverage theory: rents are higher but prices are lower when the lot is rented.)

vi) *Lift*

Whether the building has a lift surprisingly has positive effect on rental yields, out of line with the hypothesis that positive attributes have a greater effect on price than on rent. Looking at Table 12 and Table 13 we can see that this indeed is the case: the effects on both rent and price is positive, but that on rent is stronger. Once again however, though statistically significant, the economic significance to net rental yield of having a lift is limited.

vii) *Unemployment rate in area*

The town level unemployment variable has a significantly positive coefficient. If high unemployment rate in the area is on average considered undesirable, this finding is consistent with net rental yield falling with value. On the other hand, from a supply and demand perspective, this supports the hypothesis of high net rental yields in situations where the demand side is, for example, credit constrained, or may receive government subsidies to rental payments. Juntto's (2007) findings about concentration of the low income households in the rental market support this argument. Looking at Table 12 and Table 13, we can see that rents are largely unaffected by the degree of unemployment (statistically insignificant positive effect), but that prices in areas with high unemployment are lower.

viii) *Advertisement date*

All of the advertisement year dummies except for 2008 have significant t-statistics. This is very much expected, because the rental yield has historically varied quite strongly as depicted by Figure 6. As the average yields in 2008 were very close to those of 2009, it is unsurprising that 2008 is insignificant.

ix) *Metropolitan area*

All of the metropolitan area dummies have statistically significant coefficients, and all of them except for the Rest of Finland area are positive. The reason for Rest of Finland being positive is the fact that the distance from center variable measures distance from the city centers of the five metropolitan areas and as Rest of Finland is by definition far away from any of those, the yield is already inflated. Figure 7 depicts the average yields through time in the different metropolitan areas, without the control for distance from center.

x) *Constant*

The constant in the net rental yield regression is without interpretation, as it is merely one of the multipliers in equation (15). However, using the constant it is possible to obtain a predicted net rental yield by including attributes of a dwelling.

As the focus of this study is on net rental yield, we will not offer a thorough interpretation of the coefficients of annual net rent or price. However, the results of the regression analyses for annual net rent and debt-free asking price in the whole country sample are presented in Table 12 and Table 13. Additionally, a brief analysis of the key findings can be found in Table 9. Table 16 and Figure 15 present a regression analysis for net rental yield in the whole country sample, where dwelling size is split into dummy variables, but this will be discussed more thoroughly in section 0

5.5.2 Net rental yield in the HMA sample

Table 8 depicts the regression results for four OLS regressions with double-log specifications, with net rental yield as the dependent variable and the HMA sample as observations. In the table, nearly all coefficients are significant at the 1 % level (and even at the 0.1 % level), so no asterisks to indicate significance have been added. The first column of coefficients shows the first regression with no area variables, while the second column includes an the town level unemployment variable.

Table 8:

HMA sample: net rental yield

Table 8 shows regression results for the HMA sample using a double-log regression specification, following equation (14). The dependent variable is net rental yield, and the coefficients in the first column include dependent variables without area specific continuous variables. The rows with a count (n of dummy) are dummy variables. The three rightmost columns include area specific variables. As all coefficients except floor dummies, lot is rented dummy and three lift coefficients are significant at the 1 % level, asterisks marking significance have been left out. The effect of advertisement date dummies (all significant at 1 % level except for 2008, which is expectedly close to the reference class 2009) are displayed in Figure 6 and therefore marked controlled here.

Variable	n of dummy	Coefficients (t-stat in parentheses below the coefficient)			
dwelling size		-0.260	-0.252	-0.269	-0.262
		(-64.44)	(-62.66)	(-52.32)	(-64.97)
distance from center		0.011	0.004	0.011	0.009
		(5.67)	(2.31)	(5.74)	(4.82)
distance from rail stop		-0.025	-0.008	-0.026	-0.023
		(-13.92)	(-4.25)	(-14.16)	(-12.68)
distance from university		0.068	0.058	0.069	0.067
		(23.86)	(20.31)	(23.98)	(23.39)
built 1962-1983	4604	0.048	0.046	0.047	0.047
		(10.18)	(9.92)	(10.00)	(9.88)
built 1983-2009	3263	-0.060	-0.061	-0.061	-0.058
		(-10.92)	(-11.21)	(-11.08)	(-10.59)
not apartment	1280	0.142	0.142	0.142	0.179
		(19.79)	(20.00)	(19.79)	(21.66)
bottom floor	1958	0.012	0.014	0.012	0.012
		(2.46)	(2.79)	(2.44)	(2.38)
top floor	2204	-0.006	-0.006	-0.006	-0.006
		(-1.16)	(-1.38)	(-1.20)	(-1.33)
lot is rented	275	0.001	0.002	0.002	0.000
		(1.96)	(2.07)	(1.99)	(1.89)
building has lift	2511	0.013	0.013	0.013	0.014
		(2.55)	(2.25)	(2.56)	(2.61)
area average income			-0.121		
			(-19.94)		
area median ad period				0.015	
				(2.59)	
area type turnover					-0.034
					(-8.96)
advertisement date		Controlled	Controlled	Controlled	Controlled
constant		-2.163	-1.765	-2.163	-2.222
		(-132.41)	(-68.78)	(-132.45)	(-126.27)
Adjusted R-squared		0.370	0.386	0.370	0.373
n		15129	15129	15129	15129

With the HMA sample, a smaller but more consistent dataset including only one metropolitan area and thus major CBD we are able to offer further support to the hypothesis that net rental yield indeed is not constant but varies throughout the housing market, also within a single metropolitan area. The four R-squared's of the four different set-ups vary between 0.370 and 0.386, compared to 0 were the null hypothesis to hold. The individual coefficients are analyzed in the following list, but as the vast majority of the HMA coefficients are alike to those in the whole country sample, the focus will be on the differences between the results for the two samples. Two or more area variables (*vii, viii, ix*) were not included simultaneously due to strong multicollinearity, which makes the distinction of separate effects of each variable difficult.

i) *Dwelling size*

Similar effect as in whole country sample.

ii) *Distance from center / nearest rail transport stop / university*

The coefficient for distance from center is quite like the one in the whole country sample, though its effect is much weaker, which is likely to be explained by the fact that average distances within a single metropolitan area are much smaller than within the whole country.

Distance from university on the other hand has a much higher significance than distance from city center. Though somewhat surprising, as one would expect students to push rents and thus yields up near universities, a likely explanation is that as two (Schools of Economy & Helsinki University) of the four major campuses in the HMA are in the city center, the effect captured by the distance from university variable is actually more telling of distance from center.

The effect of distance from the neares rail transport stop (measured from subway and train stops) in the basic setup without area variables is somewhat unexpected: the closer the dwelling, the the higher the yield. This is in contradiction to most studies, which assume proximity to public transport to be a benefit, not a liability. However, as Laakso (1997) points out, many of these areas, especially in eastern Helsinki that are in the vicinity of subway stations, are deemed to be of low reputation. Consequently, the reason for the lower prices and rents is not the availability of public transport itself, but other possible factors that the proximity captures. This explanation seems to be partially confirmed in the second setup, which includes area average income as a zip code specific variable. After controlling for income levels, the effect of distance to rail transport stop on net rental yield practically disappears, suggesting either that either this factor is correctly priced into both rents as well as prices, or more likely (supported also by Table 14 and Table 15), it does not have an economically significant effect on either price or rent.

iii) *Building age*

Similar effect as in whole country sample.

iv) *Row or single-family house (not apartment) / dwelling floor (top/in between/bottom)*

Row or single-family house has similar effect as in whole country sample.

Bottom floor has similar effect as in whole country sample, but top floor has no statistically significant effect. Examination of its effect on rent and price (Table 14 and Table 15) shows it has no effect on those either, which could mean that in the Helsinki area, many dwellings on the top floor are in buildings without lifts. In the absence of a combined dummy (top floor and lift) it is not possible to get further insight into this.

v) *Rented or owned lot*

The dummy for rented lot is only slightly positive with very little statistical significance. This would suggest that the land leverage theory cannot explain the small variation that occurs within a single metropolitan area, but is rather suited to explain intrametropolitan variation.

vi) *Lift*

Similar effect as in whole country sample.

vii) *Average income in the area*

As mentioned before, the average income in the area has a very significantly weakening effect on net rental yields. At the same time it takes away quite a lot of the negative effect of distance from center and the positive effect of distance from rail. Similarly as for the area unemployment in the whole country sample, a likely explanation would be that in areas where average income is low, people are credit constrained and thus forced to rent even at above market prices. However the effect of area average income is much higher, even though it is only within a single metropolitan area. Looking at Table 14 and Table 15, it can be seen that area average income has a positive effect on both rents and prices, but that the effect on prices is higher. Assuming that most people consider high average income a positive trait (possibly bringing with it better services, schools, etc.), this would suggest that homeowners are more concerned about the neighborhood characteristics than tenants.

As in the HMA the areas with the highest average incomes are mostly not in the city center where the most expensive dwellings are, area average income is acts as more than a fixed effects control proxying for central location. Moreover, as dwelling price and average income do not always go hand in hand, the high significance of area average income suggests there may be more depth to the variation of rental yields than simple declining with dwelling price.

viii) Median length of rent / sale advertisement in the area

The median duration of an advertisement in the area has a slightly positive effect on net rental yield, but the effect is barely statistically significant. The variable was included as a proxy for vacancy, or inversely for liquidity. In the liquidity sense, it seems not to work very well, as lower liquidity should not decrease but increase rental yields.

The effect of vacancy rates is not as straightforward. Although a high vacancy rate always decreases effective yield (which the data used in this study does not allow us to observe) and thus requires a higher rent to compensate, a lower rent may remedy the situation if the high vacancy rates are being caused by the tenants' unwillingness to pay high asked rents.

Interestingly, looking at Table 14 and Table 15, one can note that the effects on rents and prices are much stronger than that on net rental yield, but that they cancel each other out for the most part. Somewhat surprisingly the effect is actually stronger for rents than for prices, which would suggest that vacancy rates are indeed being compensated for by higher rents. However, the possibility of reverse causality should be noted here as well. As the data used is on ask prices, it can simply be that dwellings whose ask price or rent is high take a longer time to sell or let out, or that in areas with higher average prices the decision making process of buying or renting a dwelling will on average take longer.

ix) *Turnover of dwelling type in the area*

Turnover of the dwelling type (apartment, rowhouse, single-family house) in the area has a negative effect on net rental yield. As a proxy for liquidity, it seems to function as expected: high turnover has a positive effect on both rents and prices (Table 14 and Table 15), but the effect on price is stronger, resulting in the negative effect on net rental yield. As in other asset markets, higher liquidity means lower transaction costs and thus lower liquidity premium, justifying a lower yield.

x) *Advertisement date*

Similar as in whole country sample.

xi) *Constant*

Similar as in whole country sample.

5.5.3 Summary of signs and relative magnitudes of coefficients

Table 12, Table 13, Table 14 and Table 15 in Appendix A show the results of the regression analyses for net annual rent and price for the whole country sample and the HMA sample, respectively. As they are not the focus of this study, they act merely as a sanity check for the regression analyses with net rental yield as the dependent variable. Without comparing the analysis of rents to prices, most of the signs of the regression coefficients for either one individually are as expected, and the ones that are not have been discussed in sections 5.5.1

and 5.5.2. Therefore the actual coefficients of those analyses are not discussed further. However, as it is through rent and price that net rental yield is formed, it is interesting to compare the signs and magnitudes of the coefficients. Table 9 presents a short summary of the key takeaways and possible explanations for the results of Table 12, Table 13, Table 14 and Table 15.

Table 9:

Summary of signs and relative magnitudes of regression coefficients

Table 9 summarizes key takeaways of the four regression analyses presented in Appendix A with brief interpretation. ++ means a relatively higher positive marginal effect, + a relatively lower positive marginal effect, - a relatively lower negative marginal effect and – a relatively higher negative marginal effect. 0 means that the effect is statistically insignificant. Elasticity effect as an explanation means that the coefficients for the particular variable have both had the expected sign but the magnitude of the coefficient for price has been greater. This is the case for quite many of the variables and stems from the finding that prices seem to be more elastic to both positive as well as negative attributes of the dwelling than rents.

Variable	Sample	Rent	Price	Yield	Key takeaway / possible explanations
dwelling size	Whole country	+	++	-	Elasticity effect.
	HMA	+	++	-	
distance from center	Whole country	-	--	+	Elasticity effect.
	HMA	-	--	+	
distance from rail stop					Corrected for area average income. Effect on price is negligible, while rent remains
	HMA	-	0	-	
distance from university	Whole country	-	--	+	Elasticity effect.
	HMA	-	--	+	
built 1962-1983	Whole country	-	--	+	Elasticity effect.
	HMA	-	--	+	
built 1983-2009	Whole country	+	++	-	Elasticity effect.
	HMA	+	++	-	
not apartment	Whole country	++	+	+	Elasticity effect.
	HMA	++	+	+	
bottom floor	Whole country	-	--	+	Elasticity effect.
	HMA	-	--	+	
top floor	Whole country	-	--	+	Signs for rent and price surprising, possibly bias in data. In HMA having lift offsets top floor effect.
	HMA	0	0	0	
lot is rented	Whole country	+	0	+	Results for whole country in line (though weak) with land leverage proposition. In HMA canceling effects.
	HMA	+	+	0	
building has lift	Whole country	++	+	+	Possibly renters are more likely to be in need of lift (elders, handicappeds, etc.). HMA in line with NAC.
	HMA	+	+	0	
town level unemployment	Whole country	0	-	+	Buyers more interested in area income levels than renters, possibly due to longer investment horizon.
area average income					Elasticity effect.
	HMA	+	++	-	
area median ad period					Causality may be reversed: higher rents and prices cause advertisement periods to grow longer.
	HMA	++	+	+	
area type turnover					Liquidity in area increases prices but lowers required rate of return, so rents are not increased as much.
	HMA	+	++	-	

5.5.4 Dummy regression for net rental yield in the HMA sample

Despite the benefits of the double-log regression specification, the logarithmic functional form does make some limiting assumptions about the variables that enter the regression equation, such as that all values must be strictly positive. In addition to this, it may in some cases be reasonable to argue that the effects of all the continuous variables may not indeed be continuous, but that they exhibit jumps. These discontinuities and restrictions in mind, we have performed a regression analysis where all of the variables included are dummies.

Dwelling size and distance from center in particular are variables whose effects may not be monotonous. For size, a good example is an increase in dwelling size over a threshold level that makes the dwelling large enough for one additional dweller may be valued differently than an increase that merely increases the spaciousness of the dwelling. Regarding distance from center, the same meter distance may in one case only mean a slightly longer commute while in another case, the need to start using either public trasnportation or an automobile instead of walking to the workplace.

The greatest benefit of a dummy-only regression specification is that it makes the question of model specification redundant. This comes at the cost of degrees of freedom and therefore, usability for smaller samples. In our case however, even the slightly smaller HMA sample of 15129 observations easily suffices. The HMA sample is used because, as explained in section 5.5.2, the presence of only one major CBD improves interpretability of the distance from center and the smaller variance of land values leaves less space for e.g. land leverage explaining the phenomenon.

Altogether 69 additional dummy classes were created for this regression. Dwelling size is split into 41 dummy classes as follows: 5 square meter intervals up to 100 square meters, then 10 square meter intervals up to 200 square meters and 20 square meter intervals from there onwards. The largest dwelling included was 390 square meters. Distance from center is split into 28 dummy classes, with 0.5 km intervals up to 5 kilometers, 1 km intervals from 5 kilometers to 15 kilometers and 2 kilometer intervals from there onwards. The highest distance included was slightly under 30 kilomters.

Table 10 shows the regression coefficients for the same dummy variables as in sections 5.5.1 and 5.5.2, while the coefficients for the dwelling size dummy class are plotted in Figure 13 and the coefficients for the distance from center dummy class are shown in Figure 14. The coefficients along with their respective t-stats and the sizes of each dummy class are shown in Table 11. The same analysis was done for the whole country sample as well, but as its results do not differ greatly from those in the HMA sample, they are only shown in Appendix B (Table 16 and Figure 15).

Table 10:

Dummy regression: Net rental yield in the HMA sample

Table 10 shows regression results for the HMA sample using the dummies-only regression specification. The dependent variable is net rental yield. All variables are split into dummy classes. As all coefficients except the floor dummies and the lot is rented dummy are significant at the 1 % level, asterisks marking significance have been left out. The effect of dwelling size is plotted in Figure 13, while the effect of distance from center is plotted in Figure 14. The coefficients, sizes and t-stats for both the dwelling size and distance from center dummy classes are shown in Table 11. The effect of advertisement date dummies (all significant at 1 % level except for 2008, which is expectedly close to the reference class 2009) is displayed in Figure 6.

Variable	n of dummy class	Coefficient (t-stat in parentheses below the coefficient)
dwelling size		Controlled
distance from center		Controlled
built 1962-1983	4604	0.018 (3.45)
built 1983-2009	3263	-0.091 (-15.27)
not apartment	1280	0.057 (7.16)
bottom floor	1958	0.008 (1.72)
top floor	2204	-0.008 (-1.68)
lot is rented	275	0.005 (1.73)
building has lift	2511	0.016 (2.77)
advertisement date		Controlled
constant		-4.102 (-17.12)
Adjusted R-squared		0.397
n		15129

With the dummy-only regression specification, we are again able to offer further support to the hypothesis that net rental yield indeed is not constant but varies throughout the housing market, even within a single metropolitan area. The adjusted R-squared of the dummy-only set-up is 0.397, the highest of any of the specifications. The individual coefficients are analyzed in the following list, but as the vast majority of the coefficients are alike to those in the earlier regressions, the focus will be on the differences between the results for the samples.

i) *Dwelling size*

As shown in Figure 13, the effect of larger dwelling size is quite consistently negative. The scale has been cut after the dummy class 110-120 m2, because the coefficients for dummy classes for sizes larger than that are not statistically significant even at the 10 % level, as shown in Table 11. The effect is strongest for the range of the scale, where relative increases are greatest. Overall this finding is quite well in line with the results of the double-log regressions.

ii) *Distance from center*

As shown in Figure 14, distance from center has a positive effect, but the effect is not as consistent as the one for dwelling size. Unlike the case of dwelling size, this inconsistency is not due to small amount of observations in the dummy classes (see Table 11), but more likely because of other location related variables in the HMA (smaller CBDs, coastline, etc.).

iii) *Building age*

Similar effect as in earlier regressions.

iv) *Row or single-family house (not apartment) / dwelling floor (top/in between/bottom)*

Whether the building is a row or single-family house has a similar effect as in earlier regressions.

Being on the bottom floor has similar effect as in earlier regressions, although its significance is slightly lower. Top floor is actually slightly more significant than in the earlier regressions, but again, its economic significance is limited.

v) *Rented or owned lot*

The dummy for rented lot is slightly positive but with even less statistical significance as in the HMA sample. This offers further support for the notion that the land leverage theory cannot explain the variation that occurs within a single metropolitan area, but is rather suited to explain intrametropolitan variation.

vi) *Lift*

Similar effect as in earlier regressions, though its coefficient is slightly more significant.

vii) *Constant*

Both the statistical significance as well as the coefficient for the constant are substantially smaller than they were in the double-log regressions. This is natural due to the greater amount of independent variables.

Figure 13:

HMA sample: size of dwelling dummy value plot

Figure 13 plots the effect of size of dwelling, calculated using dummies for the HMA sample. The solid line (scale on the left axis) shows the multiple with which net rental yield is multiplied for each dummy, while the dotted line (scale on the right axis) shows the net rental yield for an average dwelling (in terms of all the other parameters) in each class. Dummies with coefficients that are not significant at the 10 % level have been left out of the plot.

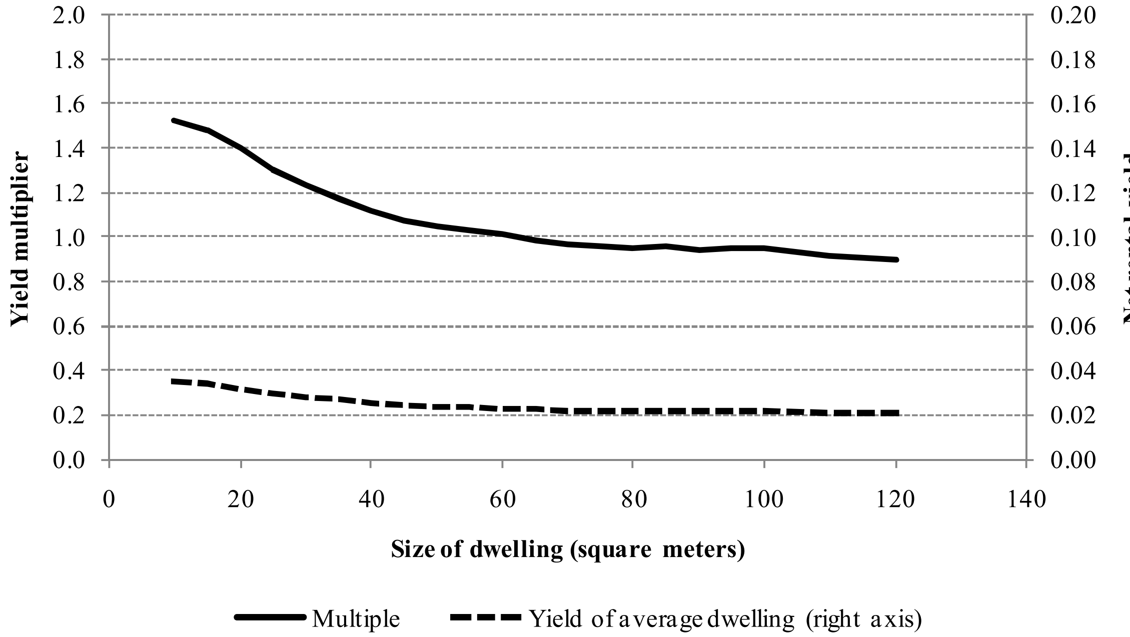


Figure 14:

HMA sample: distance from center dummy value plot

Figure 14 plots the effect of size of distance from center, calculated using dummies for the HMA sample. The solid line (scale on the left axis) shows the multiple with which net rental yield is multiplied for each dummy, while the dotted line (scale on the right axis) shows the net rental yield for an average dwelling (in terms of all the other parameters) in each class.

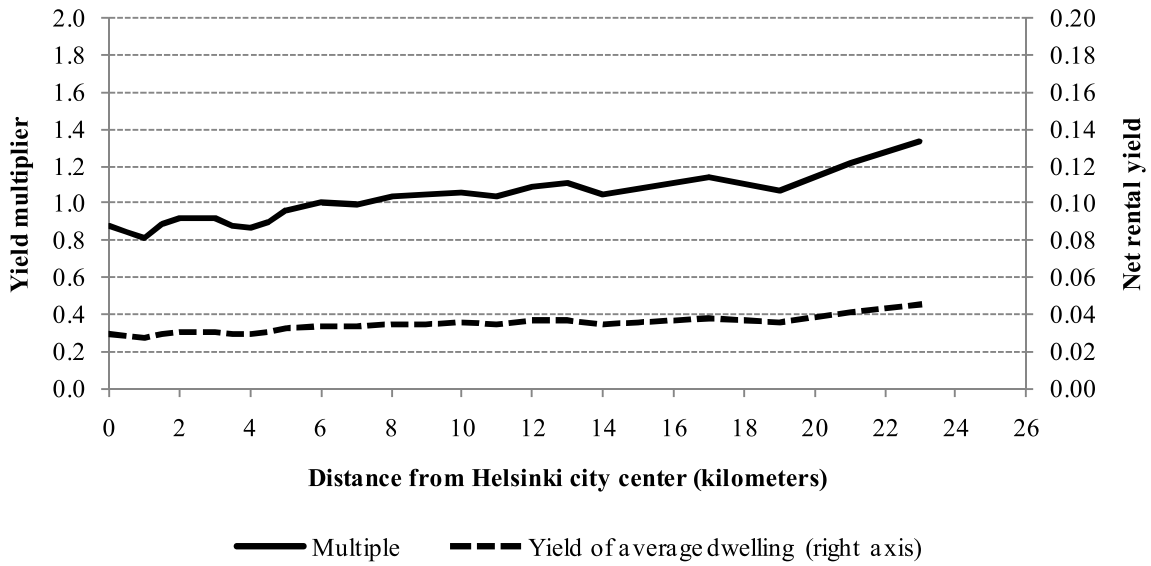


Table 11:

Dwelling size and distance from center dummy data

Table 11 shows the dummy specific counts, coefficients (converted to multipliers) and t-stats for the dwelling size and distance from center dummy classes in the dummy-only regression, for the HMA sample. The four leftmost columns show the data for the dwelling size dummy class, while the four rightmost columns show data for the distance from center dummy class. The coefficients for the rest of the independent variables are shown in Table 10 and the ones plotted here are also shown plotted in Figure 13 (dwelling size) and Figure 14 (distance from center).

Size class	n	Yield multiplier	t-stat	Distance class	n	Yield multiplier	t-stat
0 - 10 m2	0	N/A	N/A	0 - 0.5 km	233	0.88	2.78
10 - 15 m2	8	1.52	4.80	0.5 - 1 km	1014	0.85	2.61
15 - 20 m2	197	1.48	5.14	1 - 1.5 km	1727	0.82	2.44
20 - 25 m2	871	1.40	4.73	1.5 - 2 km	1147	0.89	2.85
25 - 30 m2	1335	1.30	4.23	2 - 2.5 km	1330	0.92	3.03
30 - 35 m2	1087	1.23	3.83	2.5 - 3 km	471	0.92	3.02
35 - 40 m2	912	1.17	3.43	3 - 3.5 km	597	0.92	3.03
40 - 45 m2	1223	1.12	3.13	3.5 - 4 km	432	0.87	2.76
45 - 50 m2	1524	1.08	2.85	4 - 4.5 km	370	0.86	2.72
50 - 55 m2	1588	1.04	2.62	4.5 - 5 km	284	0.89	2.88
55 - 60 m2	1318	1.03	2.50	5 - 6 km	641	0.96	3.24
60 - 65 m2	704	1.01	2.38	6 - 7 km	490	1.00	3.47
65 - 70 m2	604	0.98	2.19	7 - 8 km	537	0.99	3.42
70 - 75 m2	739	0.97	2.11	8 - 9 km	722	1.03	3.63
75 - 80 m2	671	0.96	2.01	9 - 10 km	563	1.04	3.68
80 - 85 m2	383	0.95	1.91	10 - 11 km	610	1.05	3.73
85 - 90 m2	290	0.96	1.98	11 - 12 km	824	1.04	3.67
90 - 95 m2	268	0.94	1.87	12 - 13 km	929	1.08	3.88
95 - 100 m2	250	0.95	1.94	13 - 14 km	407	1.11	3.98
100 - 110 m2	284	0.95	1.93	14 - 15 km	401	1.04	3.67
110 - 120 m2	236	0.91	1.64	15 - 17 km	775	1.07	3.83
120 - 130 m2	170	0.90	1.53	17 - 19 km	338	1.14	4.13
130 - 140 m2	117	0.87	1.32	19 - 21 km	121	1.07	3.78
140 - 150 m2	81	0.89	1.47	21 - 23 km	149	1.22	4.45
150 - 160 m2	52	0.90	1.49	23 + km	17	1.34	4.80
160 - 170 m2	60	0.89	1.49				
170 - 180 m2	41	0.87	1.26				
180 - 190 m2	29	0.92	1.62				
190 - 200 m2	20	0.89	1.44				

6 Discussion

In this paper we presented an analysis of the cross-sectional variation of net rental yields in the Finnish housing market as well as of its determinants. To conduct the analysis we used a sample of data taken from individual for rent and for sale advertisements as well as other data sources. The data was described and analyzed through descriptive statistics as well as regression analyses.

The motivation for this thesis was the hypothesis (based on individual observations and anecdotes) that rental yields are not stable across the housing market but fall as asset values increase. Though some earlier studies have documented similar findings, because of different focus areas, none of them have attempted to offer an explanation for the phenomenon. Further motivation was the relatively low amount of studies covering the housing market, especially in Finland, despite the large share that it represents of total investment assets in the economy.

The results of our study suggest that there is, as hypothesized, discrepancies between how rents and prices are set, leading to variation in rental yield. In general, the variation can be divided into two dimensions, longitudinal as well as cross-sectional, both of which seem to occur.

The main result of this study concerning cross-sectional variation is displayed in Figure 8. The observation is puzzling: net rental yield is a function of dwelling value, which falls as value increases. The economical significance of this is remarkable, as we find deviation of up to 4 % deviation between yields of the highest and lowest priced asset deciles in Table 5. This holds for dwellings of different types, sizes as well as geographies as shown by the regression analyses. Analyzing the determinants of net rental yield as well as price and rent separately in a hedonic model type regression, we find that differences in elasticities of rents and prices of the dwelling-specific characteristics explain most of the variation in yield.

There are quite expectedly a large number of determinants affecting rents and prices. The most important ones are the size of the dwelling, the distance of the dwelling from the closest CBD or other services and the age of the structure, but issues such as on which floor of an apartment building the dwelling is located, whether or not there is an elevator and how many bedrooms the dwelling has also have an effect. Naturally, area-level issues such as average income, unemployment level and population growth also have an effect. Overall, in terms of the above variables, it seems that the more attractive a dwelling is on average considered to be, the lower its expected rental yield, as it seems that tenants value dwelling attributes differently than owner occupiers.

Because the primary focus of our study was cross-sectional variation, the data we gathered is not particularly suited to examine longitudinal variation. However, both literature as well as our analysis of a shorter period, support the notion that though rent development can drift apart from that of prices on the short run, the variation is mean-reverting on a medium to long term. More specifically, it seems that rents lag behind house price development, and therefore, an increase in dwelling prices translates into a short-term fall in rental yields (and vice versa), but in the long run, the longitudinal no-arbitrage condition seems to hold, especially considering long term interest rates.

Concerning data, we find the usability of an advertisement dataset with our matching methodology good. A comparison between our data and data gathered by statistical institutes for some of the aggregate level analyses suggests that our dataset is rather closely representative of the housing market as a whole. The actual study methodologies that we have chosen, especially double-log functional form in explaining the drivers of net rental yield variation, seem to work well for this kind of analysis. The possibility of separately regressing net rent as well as price and to intuitively combine those to obtain coefficients for net rental yield allows us to better understand the factors causing the variation.

As mentioned earlier, a general explanation for the variation of both rents and prices is that rental tenants value dwelling attributes differently than owner occupiers. There can be many reasons for this, such as differences in investment horizon, cost of capital or the need to choose either one for other financial considerations such as taxation or subsidies. None of these however are sufficient from an investment perspective. If within one asset market with a fairly constant risk level, higher rental yield is attainable through better portfolio allocation, rational investors should reallocate until these abnormal yields cease to exist.

The earlier studies that have documented this phenomenon have attributed the variation simply to unobservable terms such as expected appreciation or depreciation (maintenance), but considering the size of the spread and the fact that it exists even within a single metropolitan area, this attribution is unconvincing. Nevertheless, there are two attractive explanations that could explain the falling rental yields. However, after our analysis, they too seem insufficient. Firstly, land leverage (Bostic et al., 2007), or the notion that assets with a higher share of land value built into them require a lower rate of return, was tested both by looking at an individual zip code and observing similar variation there, as well as using an owned / rented lot dummy in the regression, which was found to be economically insignificant. Secondly, the inframarginal landlord effect, proposed by Tian (2008), meaning that a large share of the higher value assets are not rented for purely economic reasons, is not consistent with the fact that similar variation exists even for the exact matched sample, where the same apartment is offered for sale and for rent and thus no emotional attachment to the dwelling should exist.

The final explanation that is left would be transaction costs: for the same euro amount invested, investing in the low-value end of the housing market requires more transactions than investing in the high-value end and thus, after taking into account transaction costs, the adjusted net rental yields are no longer different. However, considering the magnitude of the variation, it seems highly unlikely that transaction costs could account for the whole spread between the two asset classes.

After all of our failed attempts to explain the variation of rental yields in the housing market, we are left with an unexpected explanation: the housing market is inefficient, even in the weak form. Whether it is due to behavioral factors, the lack of capital to correct the situation, or simply the possibility that this variation has not been noticed by a sufficient amount of investors lies beyond the reach of this study, but as it is, there seems to be a possibility for investors in the housing market to achieve significantly better risk-adjusted returns simply through better portfolio allocation.

A conclusion such as this has many implications. As at least on a short-to-medium term, but maybe even on a longer term, assets are inconsistently priced in regard to their cash flow producing capabilities, a large portion of investors, owner-occupiers as well as renters would benefit from exploiting the variation of net rental yields. High-value owner occupiers should consider selling their low-yielding dwellings, moving into renting and buying several low-priced dwellings to pocket the spread, while tenants in low-value dwelling should attempt the opposite: move into owner-occupancy if it is financially possible. For investors, in addition to the better portfolio allocation potential, the geographical variation indicates that it is worthwhile to perform a holistic analysis of the entire country instead of focusing on one region. Policymakers should also draw conclusions, especially in terms of the way privately owned housing in Finland is subsidized: at least at the moment, housing subsidies seem to support the rental market and drive higher rents while inflow of investment capital does not drive up the price level, causing investors in the low-value end to earn consistently high rental yields.

The main contribution of this study to literature is to open up the topic of cross-sectional variation in the housing market for academic discussion. Though a few articles have indeed touched upon the subject, they have been limited to a view where, in our opinion, efficiency of the housing market has been too eagerly assumed. We feel that further discussion needs to take place with an emphasis on the investments point of view, to get better insight on the market efficiency issue.

As our study is limited to one country only, it would be interesting to see future research take upon the challenge of performing this analysis in an international setting. It would be especially interesting to see whether countries with more developed privately financed rental markets still show the same tendency for rental yields to fall as asset values increase. Additionally, as owner-occupancy is such a highly preferred tenure choice and saving method in Finland, especially for higher income households, it would be interesting to see if this same phenomenon exists in other countries where the culture is similar.

As mentioned earlier, another limitation of our study is the lack of data on which advertisements led to realized transactions, what the actual price was and whether one or both of the transaction counterparties was an institutional investor. Therefore an interesting follow-up research could be conducted by linking the advertisements to the realized prices and seeing how the rental yields are affected while at the same time being able to correct for vacancy rates. If it were further combined with data on the counterparties, more insight could be gained on the reasons behind the market inefficiency. Finally, as this data would have the benefit of reflecting real market conditions at a given moment, it could be used for other types of studies altogether, such as estimation of prices and rents using a hedonic model.

7 Bibliography

Aaron, H. "Income Taxes and Housing." *The American Economic Review* 60, no. 5 (1970).

Alonso, W. "Location and Land Use." (Harvard University Press) 1964.

Arimah, B C. "The determinants of housing tenure choice in Ibadan, Nigeria." *Urban Studies* 34, no. 1 (1997): 105-124.

Black, Fisher, and Myron Scholes. "The Effects of Dividend Yield and Dividend Policy on Common Stock Prices and Returns." *Journal of Financial Economics* 1, no. 1 (1974): 1-22.

Bostic, Raphael W, Stanley D Longhofer, and Christian L Redfearn. "Land Leverage: Decomposing Home Price Dynamics." *Real Estate Economics* 35, no. 2 (2007): 183-208.

Box, George E. P., and David Cox. "An Analysis of Transformations." *Journal of the American Statistical Association* Society Series B, no. 26 (1964): 211-252.

Capozza, D R, and G Sick. "The Risk Structure of Land Markets." *Journal of Urban Economics* 35 (1994): 297-319.

Capozza, D R, and R Helsley. "The Stochastic City." *Journal of Urban Economics* 28 (1990): 187-203.

Case, Karl E, and Robert J Shiller. "Is there a bubble in the housing market?" *Brookings Papers on Economic Activity* 34 (2003): 299-362.

Case, Karl E, and Robert J Shiller. "Prices of Single Family Homes Since 1970: New Indexes for Four Cities." *New England Economic Review*, September/October 1987: 45-56.

Case, Karl E, and Robert J Shiller. "The Efficiency of the Market for Single Family Homes." *American Economic Review* 79, no. 1 (March 1989): 125-137.

Cassel, E, and R Mendelsohn. "The Choice of Functional Forms for Hedonic Price Equations: Comment." *Journal of Urban Economics* 18, no. 2 (1985): 135-142.

Chichernea, D., N. Miller, J Fisher, Sklarz M., and White B. "A Cross-Sectional Analysis of Cap Rates by MSA." 30, no. 3 (2008).

Cropper, M L, L B Deck, and K E McConnell. "On The Choice of Functional Forms for Hedonic Price Equations." *The Review of Economics and Statistics* 70, no. 4 (November 1988): 668-675.

diPasquale, Denise, and William C Wheaton. *Urban Economics and Real Estate Markets.* Englewood Cliffs, New Jersey: Prentice Hall, 1996.

Englund, Peter, Patric H Hendershott, and Bengt Turner. "The tax reform and the housing market." *Swedish Economic Policy Review* 2, no. 2 (1995): 319-356.

Fama, Eugene F, and Kenneth R French. "Dividend Yields and Stock Market Returns." *Journal of Financial Economics* 22, no. 1 (1988): 3-25.

Finicelli, Andrea. "House price developments and fundamentals in the United States." *Bank of Italy Occasional Papers* 7 (2007).

Flavin, Marjorie, and Takashi Yamashita. "Owner-Occupied Housing and the Composition of the Household Portfolio." *The American Economic Review* 92, no. 1 (March 2002): 345-362.

Fujita, M. "Urban Economic Theory. Land Use and City Size." (Cambridge University Press) 1989.

Garner, Thesia I, and Randal Verbrugge. "Reconciling user costs and rental equivalence: Evidence from the US consumer expenditure survey." *Journal of Housing Economics* 18 (2009): 172-192.

Halvorsen, Robert, and Henry O Pollakowski. "Choice of Functional Form for Hedonic Price Equations." *Journal of Urban Economics* 10, no. 1 (1981): 37-49.

Hargreaves, Bob. "Exploring the yields on residential investment property." *Pacific Rim Real Estate Society (PRRES)*, January 2005.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. "Assessing High House Prices: Bubbles, Fundamentals, and Misperceptions." *Federal Reserve Bank of New York Staff Reports* 218 (September 2005).

Janssen, C, B. Soderberg, and Zhou J. "Robust Estimation of Hedonic Models of Price and Income for Investment Property." *Journal of Propert Investment and Finance* 19, no. 4 (2001): 342-360.

Jones, L E. "The Characteristics Model, Hedonic Prices, and the Clientel Effect." *Journal of Political Economy* 96, no. 3 (1988): 551-567.

Jorgensson, D W. "Capital Theory and Investment Behavior." *American Economic Review* 53, no. 2 (1963): 247-259.

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

Kearl, J R. "Inflation and Relative Price Distortions: The Case of Housing." *The Review of Economics and Statistics* 60, no. 4 (1978): 609-614.

Laakso, Seppo. "Public Transportation investment and residential property values in Helsinki." *Scandinavian Journal of Housing and Planning Research* 9 (1992): 217-229.

Laakso, Seppo. "Urban Housing Prices and the Demand For Housing Characteristics: A Study on Housing Prices and the Willingness to Pay for Housing Characteristics and Local Public Goods in the Helsinki Metropolitan Area." (ETLA, The Research Institute of the Finnish Economy) 1997.

Lancaster, Kevin J. "A new approach to consumer theory." *The Journal of Political Economy* 74, no. 2 (April 1966).

Maclennan, Duncan. "Some Thoughts on the Nature of Hedonic Price Functions." *Urban Studies* 14 (1977): 59-71.

Malpezzi, S. "Hedonic pricing models: a selective and applied review." In *Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*, edited by T O'Sullivan and K Gibbs. Malder, Massachusetts: Blackwell, 2003.

Mankiw, Gregory N, and David N Weil. "The Baby Boom, The Baby Bust, and the Housing Market." *Regional Science and Urban Economics* 19, no. 2 (1989): 235-258.

McCarthy, Jonathan, and Richard W Peach. "Are Home Prices the Next "Bubble"?" *Federal Reserve Bank of New York Economic Policy Review* 2, no. 2 (2004): 1-17.

Mills, Edwin S. *Urban Economics.* 5th edition. Englewood Cliffs, New Jersey: Prentice Hall, 1994.

Muth, Richard. "The Spatial Structure of The Housing Market." *Paper and Proceedings of the Regional Science Association* 7 (1961).

Oikarinen, Elias. "Studies on Housing Price Dynamics." *Turku School of Economics Series A*, no. 9 (2007).

Parsons, G R. "Hedonic prices and public goods: An argument for weighting locational attributes in hedonic regressions by lot size." *Journal of Urban Economics* 27, no. 3 (May 1990): 308-321.

Poterba, James M. "House Price Dynamics: The Role of Tax Policy and Demography." *Brookings Papers on Economic Activity* 1991, no. 2 (1991): 143-203.

Poterba, James M. "Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach." *The Quarterly Journal of Economics* 99, no. 4 (November 1984): 729-752.

Quigley, John M, and Steven Raphael. "Is Housing Unaffordable? Why Isn't It More Affordable?" *Journal of Economic Perspectives* 18, no. 1 (2004): 191-214.

Rosen, Sherwin. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82, no. 1 (January / February 1974): 34-55.

Sheppard, S. *Hedonic Analysis of Housing Markets.* Vol. 3, in *Handbook of Regional and Urban Economics*, edited by P Cheshire and E Mills, 1595-1635. Amsterdam, 1999.

Sinai, Todd, and Nicholas Souleles. "Owner-Occupied Housing as a Hedge Against Rent Risk." *The Quarterly Journal of Economics* 120, no. 2 (May 2005): 763-789.

Sirmans, G., and J. Benjamin. "Determinants of Market Rent." *Journal of Real Estate Research* 6, no. 3 (1991): 357-380.

Stiglitz, E Joseph. "Symposium on Bubbles." *Journal of Economic Perspectives* 4, no. 2 (1990): 13- 18.

Tian, Chao Yue. "Arbitrage Conditions, CAP Rates, and segmentation in the Housing Market: A Micro Study." *Working Paper, available online at: home.gwu.edu/~tianc/folder/jobmarketpaperoct2008.pdf*, 2008.

8 Appendices

APPENDIX A: NET RENT AND DEBT-FREE ASK PRICE REGRESSION RESULTS

Table 12:

Whole country sample: annual net rent

This table shows regression results for the whole country sample using a double-log regression specification, following equation (14). The dependent variable is annual net rent, and the coefficients in the first column include dependent variables without area specific continuous variables. The rows with a count (n of dummy) are dummy variables. The rightmost column includes town level unemployment, a zip code specific variable. As all coefficients except for the town level unemployment and building has lift dummies are significant at the 1 % level, asterisks marking significance have been left out. The effects of advertisement date dummies (all significant at 1 % level) and metropolitan area (all significant at 1 % level) are displayed in Figure 6 and Figure 7 respectively and therefore marked controlled here.

Variable	n of dummy	Coefficients (t-stat in parentheses below the coefficient)	
dwelling size		0.590	0.590
		(191.69)	(191.52)
distance from center		-0.070	-0.070
		(-56.10)	(-55.95)
distance from university		-0.066	-0.066
		(-48.89)	(-48.81)
built 1962-1983	10605	-0.097	-0.096
		(-29.60)	(-29.22)
built 1983-2009	7614	0.102	0.102
		(27.42)	(27.36)
not apartment	2828	0.151	0.151
		(32.55)	(32.38)
bottom floor	3829	-0.036	-0.036
		(-10.46)	(-10.45)
top floor	4152	-0.019	-0.019
		(-5.64)	(-5.63)
lot is rented	571	0.012	0.012
		(3.56)	(3.55)
building has lift	3422	0.018	0.018
		(2.42)	(2.43)
town level unemployment			0.009
			(1.16)
advertisement date		Controlled	Controlled
metropolitan area		Controlled	Controlled
constant		6.864	6.887
		(555.02)	(289.31)
Adjusted R-squared		0.796	0.796
n		27719	27719

Table 13:

Whole country sample: debt-free ask price

This table shows regression results for the whole country sample using a double-log regression specification, following equation (14). The dependent variable is debt-free ask price, and the coefficients in the first column include dependent variables without area specific continuous variables. The rows with a count (n of dummy) are dummy variables. The rightmost column includes town level unemployment, a zip code specific variable. As all coefficients except for the lot is rented dummy and building has lift dummy are significant at the 1 % level, asterisks marking significance have been left out. The effects of advertisement date dummies (all significant at 1 % level except for 2008, which is expectedly very close to the reference class 2009) and metropolitan area (all significant at 1 % level) are displayed in Figure 6 and Figure 7 respectively and therefore marked controlled here.

Variable	**n of dummy class**	**Coefficients (t-stat in parentheses below the coefficient)**	
dwelling size		0.855	0.854
		(236.77)	(236.34)
distance from center		-0.104	-0.104
		(-70.53)	(-70.71)
distance from university		-0.100	-0.100
		(-63.20)	(-63.36)
built 1962-1983	10605	-0.161	-0.163
		(-41.92)	(-42.21)
built 1983-2009	7614	0.206	0.204
		(47.47)	(46.53)
not apartment	2828	0.036	0.032
		(6.64)	(5.90)
bottom floor	3829	-0.055	-0.055
		(-13.46)	(-13.51)
top floor	4152	-0.027	-0.027
		(-6.81)	(-6.84)
lot is rented	571	-0.006	-0.005
		(-0.96)	(-0.93)
building has lift	3422	0.005	0.004
		(2.13)	(2.16)
town level unemployment			-0.042
			(-4.78)
advertisement date		Controlled	Controlled
metropolitan area		Controlled	Controlled
constant		9.005	8.891
		(620.57)	(318.42)
Adjusted R-squared		0.844	0.844
n		27719	27719

Table 14:

HMA sample: annual net rent

This table shows regression results for the HMA sample using a double-log regression specification, following equation (14). The dependent variable is annual net rent, and the coefficients in the first column include dependent variables without area specific continuous variables. The rows with a count (n of dummy) are dummy variables. The three rightmost columns include area specific variables. As all coefficients except top floor dummy and three lift coefficients are significant at the 1 % level, asterisks marking significance have been left out. The effect of advertisement date dummies (all significant at 1 % level except for 2008, which is expectedly close to the reference class 2009) are displayed in Figure 6 and therefore marked controlled here.

Variable	n of dummy	Coefficients (t-stat in parentheses below the coefficient)			
dwelling size		0.604	0.590	0.573	0.606
		(153.31)	(154.06)	(114.79)	(154.47)
distance from center		-0.035	-0.024	-0.034	-0.032
		(-18.50)	(-13.35)	(-18.27)	(-17.37)
distance from rail stop		0.018	-0.009	0.015	0.015
		(10.11)	(-4.56)	(8.16)	(8.49)
distance from university		-0.113	-0.097	-0.111	-0.111
		(-40.62)	(-35.66)	(-40.02)	(-40.08)
built 1962-1983	4604	-0.094	-0.091	-0.097	-0.092
		(-20.46)	(-20.52)	(-21.14)	(-20.09)
built 1983-2009	3263	0.089	0.090	0.085	0.086
		(16.62)	(17.44)	(15.92)	(16.20)
not apartment	1280	0.198	0.198	0.198	0.149
		(28.22)	(29.30)	(28.32)	(18.47)
bottom floor	1958	-0.025	-0.027	-0.025	-0.024
		(-5.05)	(-5.71)	(-5.17)	(-4.95)
top floor	2204	-0.004	-0.003	-0.005	-0.003
		(-0.89)	(-0.59)	(-1.03)	(-0.67)
lot is rented	275	0.018	0.016	0.020	0.019
		(3.91)	(3.90)	(4.07)	(4.02)
building has lift	2511	0.016	0.016	0.016	0.015
		(2.04)	(2.64)	(2.11)	(1.96)
area average income			0.190		
			(32.90)		
area median ad period				0.056	
				(10.17)	
area type turnover					0.044
					(12.20)
advertisement date		Controlled	Controlled	Controlled	Controlled
constant		6.818	6.191	6.817	6.897
		(428.15)	(252.90)	(429.56)	(402.84)
Adjusted R-squared		0.775	0.790	0.777	0.777
n		15129	15129	15129	15129

Table 15:

HMA sample: debt-free ask price

This table shows regression results for the HMA sample using a double-log regression specification, following equation (14). The dependent variable is annual net rent, and the coefficients in the first column include dependent variables without area specific continuous variables. The rows with a count (n of dummy) are dummy variables. The three rightmost columns include area specific variables. As all coefficients except top floor dummy, lot is rented dummy and lift dummy are significant at the 1 % level, asterisks marking significance have been left out. The effect of advertisement date dummies (all significant at 1 % level except for 2008, which is expectedly close to the reference class 2009) are displayed in Figure 6 and therefore marked controlled here.

Variable	**n of dummy**	**Coefficients (t-stat in parentheses below the coefficient)**			
dwelling size		0.864	0.842	0.841	0.868
		(216.49)	(230.44)	(166.11)	(220.48)
distance from center		-0.045	-0.029	-0.045	-0.042
		(-24.04)	(-16.59)	(-23.86)	(-22.30)
distance from rail stop		0.043	0.000	0.040	0.038
		(24.10)	(-0.07)	(22.42)	(21.48)
distance from university		-0.181	-0.156	-0.180	-0.178
		(-64.23)	(-59.72)	(-63.73)	(-63.91)
built 1962-1983	4604	-0.142	-0.137	-0.144	-0.138
		(-30.48)	(-32.40)	(-30.94)	(-30.13)
built 1983-2009	3263	0.149	0.151	0.146	0.144
		(27.49)	(30.65)	(26.95)	(27.03)
not apartment	1280	0.055	0.056	0.055	-0.031
		(7.80)	(8.73)	(7.82)	(-3.79)
bottom floor	1958	-0.037	-0.041	-0.037	-0.036
		(-7.46)	(-9.04)	(-7.55)	(-7.37)
top floor	2204	0.001	0.004	0.001	0.003
		(0.32)	(0.91)	(0.22)	(0.71)
lot is rented	275	0.017	0.014	0.018	0.019
		(1.91)	(1.85)	(2.02)	(2.10)
building has lift	2511	0.003	0.003	0.003	0.000
		(0.56)	(0.30)	(0.51)	(0.71)
area average income			0.311		
			(56.40)		
area median ad period				0.042	
				(7.38)	
area type turnover					0.078
					(21.32)
advertisement date		Controlled	Controlled	Controlled	Controlled
constant		8.981	7.957	8.981	9.120
		(556.57)	(340.91)	(557.53)	(530.89)
Adjusted R-squared		0.864	0.887	0.864	0.868
n		15129	15129	15129	15129

APPENDIX B: WHOLE COUNTRY SAMPLE REGRESSION WITH SIZE AS DUMMIES

Table 16:

Dummy regression: Net rental yield in the Whole country sample

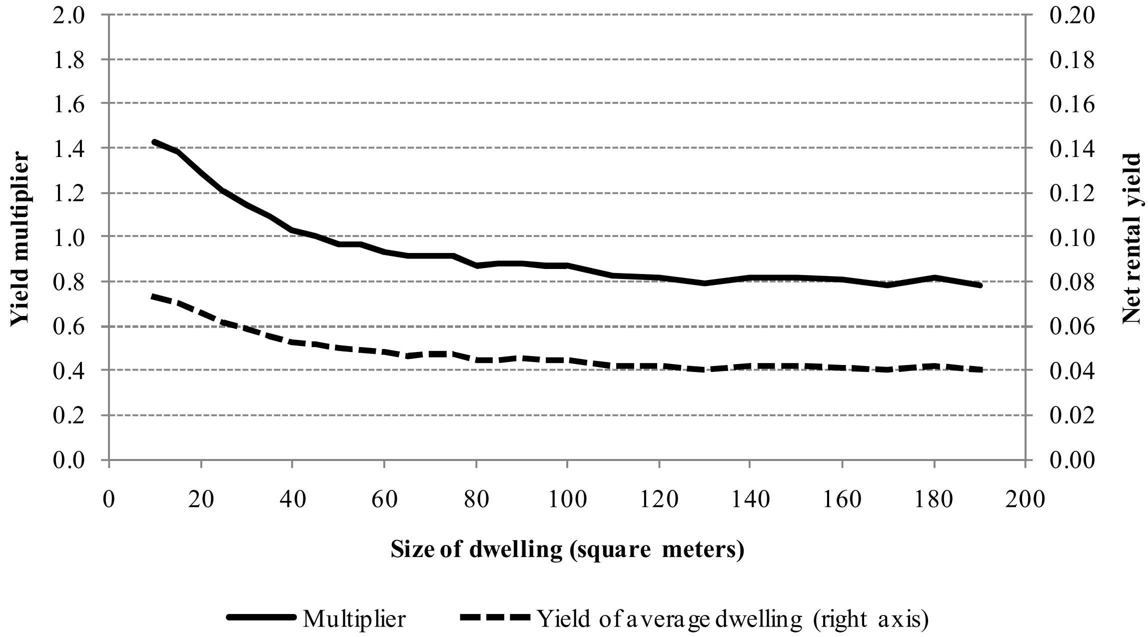
Table 16 shows regression results for the whole sample using the dummies-only regression specification. The dependent variable is net rental yield. All variables are split into dummy classes. As all coefficients except the top floor dummy are significant at the 1 % level, asterisks marking significance have been left out. The effect of dwelling size is plotted in Figure 15. The coefficients, sizes and t-stats for the dwelling size dummy class are omitted, but their magnitudes, as shown in Figure 15 are quite similar as in Table 11 for the HMA sample. The effect of advertisement date dummies (all significant at 1 % level except for 2008, which is expectedly close to the reference class 2009) is displayed in Figure 6.

Variable	n of dummy class	Coefficient (t-stat in parentheses below the coefficient)
distance from center		0.035
		(25.04)
distance from university		0.034
		(23.12)
built 1962-1983	10605	0.066
		(18.14)
built 1983-2009	7614	-0.099
		(-24.23)
not apartment	2828	0.087
		(15.61)
bottom floor	3829	0.020
		(5.11)
top floor	4152	0.009
		(2.45)
lot is rented	571	0.016
		(3.89)
building has lift	3422	0.013
		(4.40)
advertisement date		Controlled
metropolitan area		Controlled
constant		-3.794
		(-18.01)
Adjusted R-squared		0.380
n		27719

Figure 15:

Whole country sample: size of dwelling dummy value plot

Figure 15 plots the effect of size of dwelling, calculated using dummies for the whole country sample. The solid line (scale on the left axis) shows the multiple with which net rental yield is multiplied for each dummy, while the dotted line (scale on the right axis) shows the net rental yield for an average dwelling (in terms of all the other parameters) in each class. Dummies with coefficients that are not significant at the 10 % level have been left out of the plot.



APPENDIX C: CORRELATION TABLES

Table 17:

Correlation table for the whole country sample

Whole country sample	built 1962-1983	built 1983-2009	not apartment	bottom floor	top floor	distance from center	distance from university	dwelling size	town level unemployment	net annual rent	sales price	net rental yield	lot is rented
built 1983-2009	-0.4844	1											
not apartment	-0.1231	0.3528	1										
bottom floor	0.0484	-0.0259	-0.1349	1									
top floor	0.0665	-0.0853	-0.1415	-0.168	1								
distance from center	0.2644	0.1534	0.1691	0.0621	0.048	1							
distance from university	0.2747	0.166	0.1836	0.052	0.0413	0.7933	1						
dwelling size	-0.0015	0.306	0.4349	-0.0644	-0.042	0.1343	0.1403	1					
town level unemployment	0.0625	0.0164	-0.1415	0.0303	0.0232	0.0857	0.1074	-0.0489	1				
net annual rent	-0.3526	0.261	0.3516	-0.1181	-0.079	-0.2637	-0.2853	0.5976	-0.4026	1			
sales price	-0.4016	0.298	0.2922	-0.1183	-0.0768	-0.3186	-0.341	0.6158	-0.3772	0.8973	1		
net rental yield	0.2977	-0.2218	-0.0652	0.0651	0.0387	0.2616	0.275	-0.366	0.1658	-0.3272	-0.7107	1	
lot is rented	0.0113	0.0228	0.0367	-0.0073	-0.0018	0.0343	0.0414	0.0267	0.0077	0.0196	0.0025	0.0258	1
building has lift	-0.0075	-0.0396	-0.0888	-0.0444	-0.0377	-0.0319	-0.0371	-0.0673	-0.114	0.013	-0.0061	0.0338	0.0853

Table 18:

Correlation table for the HMA sample

HMA sample	built 1962-1983	built 1983-2009	not apartment	bottom floor	top floor	distance from center	distance from rail stop	distance from university	dwelling size	area average income	area type turnover	area median ad period	net annual rent	sales price	net rental yield	lot is rented
built 1983-2009	-0.3468	1														
not apartment	-0.0416	0.3707	1													
bottom floor	0.0463	-0.027	-0.1172	1												
top floor	0.0404	-0.0803	-0.1255	-0.1592	1											
distance from center	0.3325	0.2931	0.2479	0.0533	0.0146	1										
distance from rail stop	0.1011	0.1029	0.2214	0.0386	-0.0286	0.269	1									
distance from university	0.3893	0.3462	0.282	0.0589	0.0163	0.7653	0.2038	1								
dwelling size	0.0368	0.3377	0.4675	-0.0731	-0.0416	0.1692	0.2003	0.1677	1							
area average income	-0.1477	-0.0764	0.009	-0.0057	-0.0343	-0.301	0.3229	-0.3376	0.1072	1						
area type turnover	-0.1415	0.2067	0.5459	-0.0809	-0.0969	-0.0174	0.1821	0.0049	0.2258	0.3072	1					
area median ad period	0.0346	0.248	0.3296	-0.042	-0.0274	0.098	0.258	0.0843	0.6921	0.1981	0.1765	1				
net annual rent	-0.259	0.3114	0.4248	-0.1104	-0.0646	-0.1598	0.1295	-0.1972	0.7569	0.3267	0.3222	0.5707	1			
sales price	-0.3007	0.3086	0.3506	-0.1107	-0.0525	-0.2124	0.1428	-0.2699	0.7807	0.3932	0.3054	0.5811	0.8906	1		
net rental yield	0.2159	-0.1471	-0.0515	0.055	0.0061	0.19	-0.0919	0.2512	-0.4228	-0.3019	-0.1228	-0.303	-0.2596	-0.6702	1	
lot is rented	0.0079	0.0417	0.0386	0.0021	0.0027	0.0264	0.0035	0.0324	0.0275	-0.0085	0.0117	0.0074	0.0363	0.0221	0.0125	1
building has lift	0.0015	-0.0119	-0.1024	-0.0529	-0.0493	-0.0566	-0.0363	-0.0498	-0.0665	-0.0032	-0.0347	-0.0513	-0.0794	-0.085	0.0509	0.0856