Modelling existing home sales in densely and thinly populated areas in the Netherlands

Abstract

This study investigated which determinants explain the number of sold homes in the Netherlands between 2000-2013. To provide more depth to this research, differences in determinants between densely and thinly populated areas were investigated. Given that this research takes into account those differences, data on municipality level has been used.

To find an answer to this question, this study utilized two types of models: one to estimate the turnover-rate (number of sold homes divided by the owner occupied housing stock) in levels and one in first differences. Amongst others, this study finds that variations in the turnover-rate in levels is explained well by municipality demographic and municipality economic variables. This study finds key determinants of the turnover-rate stated in first differences to be (multiple lags of) the turnover-rate, house prices, mortgage interest rates and household income. Regressions on urbanization level sheds light on the differences between urbanization levels. Lagged turnover-rates and lagged mortgage interest rates are found to influence High-/very-high urban areas more compared to Non- and low- urban areas.

Keywords: Turnover-rate model, Number of sold homes, transactions, transaction frequency, House price, housing market, the Netherlands, municipality, urbanization.

Master Thesis MSc Business Economics- Finance & Real Estate Finance

University of Amsterdam

Lisa Hoving (10177272)

Supervisor: Prof. dr. M.K. Francke

February-July 2016

Statement of Originality

This document is written by Student Lisa Hoving, who declares to take full responsibility for the contents of this document.

I declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it.

The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

Acknowledgements
I would like to take a moment to thank everyone that helped me during the process of writhing this thesis. First of all, I would like to thank again my supervisor Marc Francke, who helped me greatly throughout the whole process of writing this thesis. Secondly, I would like to thank Martijn Dröes for his help and expertise with putting my model together. Lastly, I would like to thank Paul de Vries and Pieter van Dalen for their support and guidance. Everyone put so much effort into helping me write this thesis, and I could never thank all of you enough for your kindness. I feel extremely lucky to have written my thesis at the Rabobank, and enjoyed every second of it.

Of course, this section would not be finished if I did not thank all my family and friends. You know how much I love you.

Content

1. Introduction

The number of existing home sales together with house prices indicate housing market conditions (Fisher, Gatzlaff, Geltner & Haurin, 2004). House prices alone are not sufficient to convey the complete story. As explained by de Wit, Englund and Francke (2013): House price fluctuations itself do not explain booms and busts of housing markets. In economic upturns (downturns), housing markets are also more (less) liquid.

With regard to house prices, there have been plenty studies looking for its determinants, like DiPasquale and Wheaton (1994), Francke (2010) or de Vries and Boelhouwer (2009). Research focusing on the number of sold houses mostly investigates the link between house prices and the number of transactions[1], but neglects to investigate what actually determines the number of sold homes.

The limited research done on determinants of the number of existing home sales is mainly done on a national- or even international base. Studies on a regional level are scarce. With regard to explaining the determinants on the number of home transactions, there is no research done on municipalities to date in The Netherlands. This is surprising, as municipalities in the Netherlands differ substantially. Not only in the number of sold homes (Oevering, 2014, p91), but also in economic and demographic variables like for example household income and the number of children being born (see Figure 1 and 2). Given those differences, it is likely that those municipalities will also react differently to different shocks.

To get a comprehensive understanding of what determines the number of existing home sales, research needs to take into account differences between municipalities. This research fills this gap by taking municipality differences into account in the Netherlands. A way to start investigating differences between municipalities could be by looking at differences between thinly and densely populated areas. Therefore my research question is: ***which determinants explain the number of houses sold in the Netherlands between 2000-2013, and are there differences between densely and thinly populated areas?***

Explaining determinants of the number of sold homes within municipalities is important for several parties. For families selling a house, seeing the housing market in their area freeze up could mean that time on the market will increase for their house. This happened for example in the 2008 financial crisis, where there were quite a lot of homes on the market, but not enough buyers. As a result, many households saw the time on market increase. Even now when the crisis has mostly passed, the number of existing home sales still differs substantially from one municipality to another. This is even the case when you correct for the size of the housing stock. To know why this is the case and potentially help

[1] Stein, 1995; Follain & Velz, 1995; Genesove & Mayer, 2001; Sinai & Souleles, 2015.

households selling their home, more research needs to be done. Forecasting the number of existing home sales is also important for financial institutions such as banks, insurance companies and pension funds. To get a comprehensive view about mortgage production and the associated risks, the number of home sales need to be taken into account. Municipalities differ greatly, not only amongst each other but also over time. This is also what has been observed: some municipalities have been recovering faster after the 2008 crisis than others (Aalders & van Dalen, 2016). Given the stated differences in municipalities, national models are insufficient in explaining movements in the number of home sales for specific municipalities.

It has to be mentioned that this study utilized the turnover-rate as a measure of the number of existing homes sold. The turnover-rate was computed by dividing the number of existing homes sold by the owner occupied housing stock. This made municipalities more comparable, as some municipalities have a higher number of home sales just due to their higher housing stock. In addition, this research focused solely on existing homes. This means that newly constructed homes were taken out of the picture.

In order to find an answer to the research question, this paper took a two sided approach. Firstly, the turnover-rate in stated levels is regressed from year to year on possible determinants using simple OLS. Then, the turnover-rate stated in first differences is regressed using a panel autoregressive distributed lag(1) model (ADL). In addition to this model, a bivariate panel vector autoregressive model (PVAR) (1) has been used. Results of the latter are not directly reported but can be found in **Appendix G.**

This study finds that that young people aged 15-44, the number of children being born, the average amount of people within one household and income are key determinants for estimating the turnover-rate stated in levels for the Netherlands between 2000-2013. It is found that young people aged 15-44 tend to move more often. The number of children being born within a municipality significantly increased the turnover-rate. Household income has become increasingly significant over time for the decision to move, affecting the turnover-rate positively. However, it is also found that more people within one household might decrease the average turnover-rate.

On the other hand, this study finds that changes in house prices, lagged turnover-rates, mortgage interest rates and household income are key determinants for estimating the turnover-rate stated in first differences in the Netherlands between 2000-2013. House prices are found to have a small negative effect on the turnover-rate. In addition, the lag of the turnover-rate is found to affect the turnover-rate in a positive manner. This could indicate that turnover-rates tend to keep on increasing (decreasing). Interest rates appear to have a long lasting negative effect on the turnover-rate. An increase in income is found to have a small positive effect on the average turnover-rate. This study also found evidence for differences between urbanization types. Lagged turnover-rates and lagged mortgage interest rates are found to influence High-/very-high urban areas more compared to Non- and low- urban areas.

The structure of this thesis is as follows: Section two covers the existing literature. From this, hypotheses have been derived. Section 3 explains the methodology. Section 4 discusses data description, descriptive statistics and preliminary tests. Section 5 elaborates on the empirical results. Lastly, section 6 covers the conclusion.

2. Literature Review

Housing market literature generally consists of decomposing the relationship between house prices and the number of home sales. Research focusing solely on the determinants of the number of sold homes is rare, mainly because most studies finds both to be intertwined, simultaneously affecting each other[2]. Therefore, first some theory behind the link between home sales and house prices was explained in the next section. Section 2.1.1 focusses on the link between home sales and house prices for the Netherlands specifically. As house prices and the number of home sales are likely to be related, section 2.1.2 elaborates briefly on the determinants of house prices. Section 2.2 broadens the scope by discussing literature that focusses on other determinants of the number of sold homes than just price. All determinants used in these studies are summarized into Table 1. Section 2.3 elaborates on the importance of focusing on differences within municipalities when explaining the number of excising home sales.

Please note that some studies choose to use the turnover-rate instead of the number of home sales. The turnover-rate is generally defined as the number of home sales divided by the (owner occupied) housing stock. When a study utilizes this definition, it was mentioned in the literature review.

2.1 The Link Between House Prices and the number of sold homes

As described by Dröes and Francke (2016), most research explains the relationship between house prices and the number of sold houses in one of three ways: by looking at either (1) credit constraints (Stein, 1995; Follain & Velz 1995) (2) nominal loss aversion (Genesove & Mayer, 2001), (3) hedging incentives (Sinai & Souleles, 2005). As explained by de Wit et al. (2013) and Francke and van Dijk (2015), another way to explain this relationship is by looking at search and matching models.

Stein (1995) investigates the link between house prices and the number of sold homes in the US, and finds a positive relationship. He explains this by looking at credit constraints, explaining that households need downpayment on their house. If house prices increase, this downpayment is relatively easily paid for second time buyers. However, if house prices are decreasing, this might restrict families from moving, in effect also decreasing the number of sold homes.

Follain and Velz (1995) also ascribe the link between the number of sold houses and house prices to credit constraints. In their research, they composed a structural model of the housing market. This includes a supply- and demand function of residential real estate in the

[2]Amongst others: Dröes and Francke, 2015; Stein, 1995; Follain & Velz, 1995; Genesove & Mayer, 2001; Sinai & Souleles, 2015.

US. In addition, they also includes an equation to explain the number of households. Follain and Velz (1995) distinguish themselves from others by including an equation that explains the turnover-rate. This estimated turnover-rate is then included in the supply equation of the model. Against their expectations, they find house prices and volume of sales to be negatively related. Follain and Velz (1995) assign this negative effect to the reduced importance of downpayment and liquidity constraints in the 1980-1990s.

Genesove and Mayer (2001) take a more behavioral approach, and try to explain the link between house prices and the number of sold homes by looking at nominal loss aversion in the Boston condominium market. They explain that home owners in a down market do not want to realize nominal losses. Therefore they set their prices higher than you would expect in such a market. This results in a longer time to market, but also a higher price for their house when they eventually sell. This would mean that in a down market, the relationship between house prices and the number of sold homes is positive. This however does not have to be the case in when the housing market is experiencing a boom.

Sinai and Souleles (2005) explain the link by looking at hedging incentives in the US. They explain everyone needs to live somewhere. Hence, everyone is exposed to housing risk. This makes home-ownership a tradeoff between interest rate risk (when renting) and house price risk (when owning a home). If volatility of interest rate increases, the demand of houses and the number of sold houses increases (Sinai & Souleles, 2015). All in all, they find that the risk of owning a home declines with a persons expected horizon. The risk of owning a home also declines with the correlation of housing costs between current and future home locations.

The main theory behind search and matching models is that the housing market is characterized by not having central exchange. Buyers and sellers will have to look for each other until there is a match. The house is sold only when the reservation price of the seller is lower than the reservation price of the buyer. This matters for the relationship between house prices and the number of sold homes. For example, Berkovec and Goodman (1996) develop a search and matching model for the housing market. Amongst others, they find that the number of excising homes responds faster to changes in demand shocks on the housing market than do house prices. Berkovec and Goodman (1996) suggest that the turnover-rate thus might be a better measure of high frequency changes in housing demand then house prices. Genesove and Han (2012) also utilize a search and matching model. They find that the time on market for both buyers and sellers and the number of homes they will visit will decrease when demand decreases. Genesove and Han (2012) conclude that this is consistent with a search and matching model where sellers respond to demand shocks with a lag. Both papers suggest that the correlation between house prices and the number of sold homes (or the turnover-rate) is positive.

2.1.1 Research on Dutch home Sales

For the Netherlands specifically, the only study investigating the correlation between house prices and the number of sold homes has been done by de Wit et al. (2013). They used a timespan between January 1985-December 2007. As a measure of the number of sold homes, they use the turnover-rate (which they call “rate of sale”). To find the link between house prices and the rate of sale, they estimate a Vector Error Correction-Model (VECM) for the Netherlands stated in first differences. All in all, they find a strong positive correlation between house prices and the rate of sale. However, their results do not enable them to find evidence for one of the three mentioned effects specifically. Their findings do suggest that the found positive correlation stems from the correlation of price and the turnover-rate, not the rate of entry. This means that, for The Netherlands, they do not find evidence for the credit constraint theory specifically. Another important finding of this paper is the way information is incorporated in prices and the turnover-rate. For the Netherlands, de Wit et al. (2013) find evidence for a gradual adjustment of expectations when new information about market fundamental arises. This means that the turnover-rate immediately increases (decreases), but decreases (increases) again after. The change in house prices is gradually but permanent. Other research like Andrew and Meen (2003a), Berkovec and Goodman (1996) and Hort (2000) find similar results.

Based on the above, the following hypothesis has been derived:
H1: The turnover-rate and house prices are significantly related.

2.1.2 Determinants of House Prices

DiPasquale and Wheaton (1994) investigated determinants of transaction prices by looking at the United States between 1960-1990. They developed a stock-flow model to look at housing demand and supply. They find that, amongst others, household income, the homeownership rate and rents are important in explaining house prices. Overall, this model explained prices quite well.

In order to model house prices for the Netherlands between 1970-2009, Francke (2010) estimates an error correction (ECM) model. In this paper, Francke (2010) builds on an earlier model by Francke, Vujić, and Vos (2010). This ECM model consists of only demand factors. User costs (interest rates) as a percentage of house prices, income per household, and financial capital explain the long term variation in price well. Supply factors like construction, housing stock, and construction costs, cannot explain long term house price variations well. Francke (2010) explains this by stating that the housing market is a stock market. In the short run, supply does not respond to demand shocks. In the medium-long term this might also be the case, mainly due to government interference. For explaining house prices in the short term, Francke et al. (2010) uses yearly changes in user costs, financial capital per household and GDP growth.

Carrillo, de Wit and Larson (2015) also investigate transaction prices. They try to predict house price appreciation by variables that measure market tightness. The reasoning behind this is those so called tight markets might be more liquid, more expensive and experience higher turnover-rates. By doing so, they look at seller's bargaining power and sale probability. To find evidence for their hypothesis, data on the Netherlands including 36 regions are used. To this dataset Carrillo et al. (2015) add 13 medium and large MSAs in the United States and 41 ZIP Codes from Fairfax county, Virginia. In their study, both autoregressive distributed lag (AD) models and Vector Autoregressive (VAR) models were used. Their findings suggest that the sale probability and sellers bargaining power can predict home price appreciation.

Francke and van Dijk (2015) used internet search data to predict house prices. In contrast and to Carrillo et al. (2015) they measure market tightness by the amount of internet searches on Funda.nl. Reasoning behind this is that people will start looking for a home on the internet, hence the amount of Funda.nl clicks measures market demand. Liquidity is measured by the *rate of sale*, which is defined as the number of house sales divided by the number of houses for sale. This measure is thus similar to the turnover-rate, with the difference that here the denominator is not housing stock, but number of houses for sale. House prices are measured by a house price index. Their main findings consist of three things (1) market tightness affects liquidity positively, (2) market tightness affects prices positively and Granger cause changes in house prices, (3) liquidity responds fast to shocks in demand factors and is temporary, meanwhile house prices respond gradually. However, on the contrary to liquidity, changes in house prices are permanent.

2.2 Determinants of Existing home sales

In order to find a way to predict the number of home sales, Dua and Miller (1996) investigates Connecticut home sales. They used a Bayesian Vector Autoregressive (BVAR) and vector Autoregressive (VAR) method, including an index based on the unemployment rate and building permits to measure economic conditions. They also included house permits, house prices, mortgage rates and buyer attitudes. Data on buying attitudes came from household responses on a survey of the University of Michigan. The question answered was 'Generally speaking, do you think now is a good or bad time to buy a house?'.
The index was subsequently created as follows:

$$Index = 100 * \ (good + 0.5 \ * uncertain)/(bad + good + uncertain).$$

Dua and Miller (1996) find that including the buyer attitudes do not increase model accuracy when the other mentioned variables are added as well. Overall, the BVAR method provided the most accurate forecasts.

On the same topic, Dua and Smyth (1995) composed a model for US as a whole. They included personal disposable income, the unemployment rate, house prices, the mortgage rate and buying attitudes. For the US they find the same results as for Connecticut: Including buyer attitudes together with the other variables do not add to the explanatory value of the model.

Fisher et al. (2004) also looked for determinants that can explain the number of sales. This study distinguishes itself because it investigates commercial properties, not houses. Fisher et al. (2004) used a likelihood of sale (probit) model. This model has 3 types of independent variables: Market conditions, property and locational characteristics[3]. Amongst others, they include variables on employment and transaction prices as a proxy for market conditions. Fisher et al. (2004) hypothesizes that the number of sales is often found to be pro cyclical. Hence they expect an increase in employment and transaction prices to go together with an increase in the number of sold commercial properties. For both variables they find a positive relationship with the number of sold commercial properties.

Clayton, Miller, and Peng (2010) utilized a bivariate panel vector autoregressive (PVAR) model. As dependent variable they used the turnover-rate. Clayton et al. (2010) uses quarterly data on American Metropolitan Statistical Areas (MSA) between 1990:2-2000:2. Their findings suggests that the housing market is affected by three markets: (1) the labor market, (2) the mortgage market and (3) the stock market. The labor market is measured by income, employment and unemployment. Clayton et al. (2010) finds that both the employment rate and income have a significant positive effect on the turnover-rate. They explain that both income and employment might increase housing demand, hence increasing the turnover-rate. In addition, unemployment has a significant negative effect on the turnover-rate. Clayton et al. (2010) ascribes these findings to the lock in phenomenon, where families hit by an increasing unemployment rate experience financial constraints and need to raise housing prices (and hence less houses are sold). The mortgage market is measured by the mortgage rate and the trend of the mortgage rate. Those variables are found to be respectively negatively and positively related to the turnover rate. Clayton et al. (2010) explains that these results are in line with rational behavior. When mortgage rates are high (low), house prices and the number of sold homes are low (high). However, when mortgage interest rates increase, homebuyers are best off to buy soon. If they are decreasing, postponing would be the best thing to do. Lastly, the stock market is measured by the S&P500 and the trend of the S&P500 index. Surprisingly, the trend is significantly negative in the turnover-equation. Clayton et al. (2010) does not provide a direct explanation for this

[3] Property and locational characteristics are of less importance for my study, as they are property characteristics of commercial properties.

phenomenon, but suggests that private valuations of sellers might be influenced by the expectation of the economy. In addition to looking at those three markets, Clayton et al. (2010) also looked at the relationship between house prices and the turnover-rate. They find that house prices Granger cause trading volume in a one sided manner: only decreases in house prices lead to less sold homes. Trading volume does Granger cause prices, but only when the market experiences low supply elasticity. They ascribe this effect to the credit constrained theory as explained by Stein (1995) and loss aversion, which are both most prominent in markets with decreasing house prices.

Dröes and Francke (2016) take a more elaborate view by looking at the link between house prices and the turnover-rate in Europe. They used a reduced form bivariate panel vector autoregressive PVAR(1) model to test their hypotheses. One of the dependent variables is turnover, the other one is house prices. They find interest rates and GDP both to be key determinants in explaining the turnover-rate and house prices, finding a negative and positive relationship respectively. They explain this effect by saying that low interest rates and high income both makes obtaining a mortgage easier, hence resulting in higher turnover-rate estimates. This effect is greater in the turnover-rate equation than in the house price equation, suggesting that this effect mainly goes through the turnover-rate. In addition to this, they included the lag of both the turnover-rate and transaction prices in their model, finding significant positive coefficients for both. Dröes and Francke (2016) conclude that both turnover transaction prices have significant momentum. They explain this effect by stating that when the housing market is increasing, it tends to do so for several periods. Including outstanding mortgage balance to GDP, population, the share of young population (age 18-30) and the Harmonized Index of Consumer Prices (HCIP) did not result in more explanatory power within the model. The lack of any effect for the share of young population is explained by the fact that the model is stated in first differences, the share of young population likely did not change enough to find significant estimates. Even though Dröes and Francke (2016) do not find significant estimates, they state that both determinants are likely to explain variances in the price-turnover relationship. As the share of young population generally has less accumulated wealth, they will likely respond more heavily house price shocks. Meanwhile, the older population has more accumulated wealth and will likely move to smaller homes. For them, house price changes does not matter as much. One of the takeaways from this research is that the feedback between prices and turnover cannot be ignored. Dröes and Francke (2016) even find that an increase of one percent in lagged prices decreases turnover by 0.74%, meanwhile one percent increase in the turnover-rate increases house prices by 0.24%. Not taking account of this feedback leads to large bias of coefficients in both the price and turnover-rate.

Based on the existing literature, hypothesis 2-5 are as follows:

H2: If employment increases, the turnover-rate will be positively affected.

H3: An increase in income has a positive effect on the turnover-rate.

H4: An increase in mortgage rates has a negative effect on the turnover-rate.

H5: The turnover-rate and house prices are subject to momentum.

Table 1: Used variables in papers discussed in section 2.2

	Dua and Miller (1996)	Dua and Smyth (1995)	Fisher et. al (2004)	Clayton, Miller, and Peng (2010)	Dröes and Francke (2016)
Dependent Variable:					
Number of sold homes	X	X			
Turnover-rate stated in first differences				X	X
Dummy variable (1 if house is sold 0 otherwise)	X		X		
Explanatory variables:					
Economic					
One or more lag(s) of dependent variable				X	X
House Prices (sale price, hedonic price, index or otherwise)		X	X	X	X
Mortgage rate	X	X	X	X	X
Real mortgage rate					X
Economic activity index (Coincident/Leading index)	X				
Unemployment	X	X			
Employment			X	X	
Real Income	X	X			X
Income				X	X
Buyer attitude	X	X			
Building permits authorized	X				
Stock index			X		X
Yield on treasury notes			X		
Mortgage balance to GDP					X
Inflation					X
Demographic					
Age					X
Population					X
Other					
Square footage			X		
Age Property			X		

2.3 Regional Differences

Looking at national housing markets, one has to keep in mind that they consist of smaller markets. Andrew and Meen (2003a; 2003b) stresses this importance in their two part paper. The first article explains macroeconomic influences on the number of sold homes in Britain. Even though they find a positive relationship between house prices and the number of sold homes, it provides little guidance to underlying causes. In their second study, they find those underlying causes to be in demographic differences between regions. Furthermore, Meen (1999) investigated the ripple effect in Britain: the tendency of house prices to increase in the south-east first and progressively spread out to the rest of Britain over time. He states that regional house prices can be explained in three ways: (1) variables equal to all regions; (2) differences in economic growth within regions; (3) Structural demographic differences in regional housing markets. Again, the conclusion is the same: taking account of only variables equal to all regions is not enough. One has to take into account regional variables as well. Overall, these papers stress the importance of looking at microeconomic and regional factors. Macroeconomic factors alone cannot explain the variation of the housing market on a lower scale.

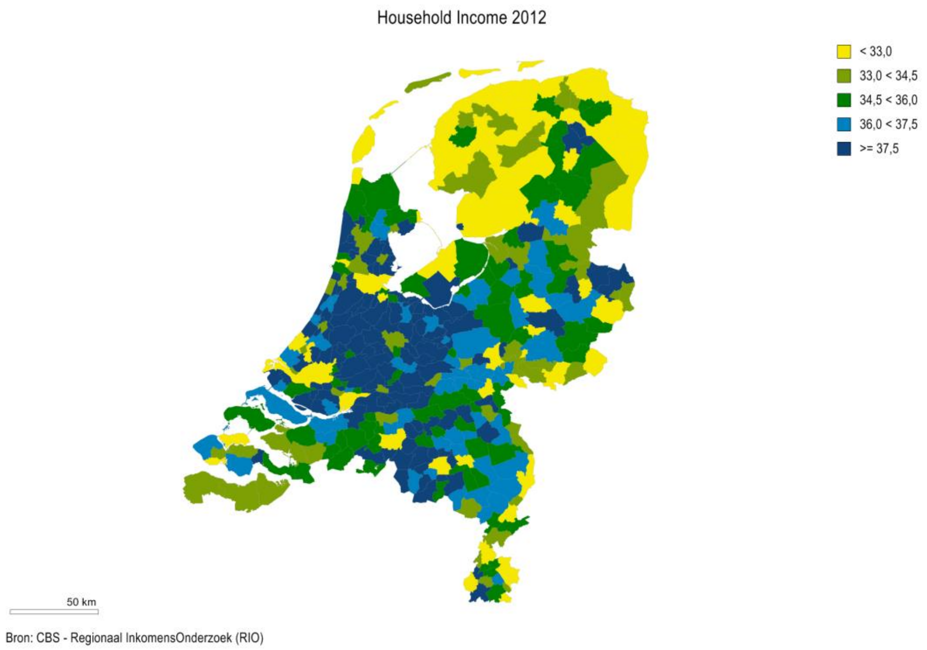
Even though Andrew and Meen (2003a; 2003b) and Meen (1999) investigated Britain, it is likely that their findings are applicable for the Netherlands as well. In the Netherlands too, municipalities differ greatly amongst each other, both economically and demographically. For example household income is lower in northern regions (around 33 thousand) compared to the Randstad region (>37.5 thousand, see Figure 1). Demographic variables also show considerable differences throughout municipalities. For example the number of children being born differs greatly. Most children are being born in large cities, meanwhile the least are being born in small towns (see Figure 2). Given those differences, it is likely that macroeconomic factors alone will not explain all the deviation in the number of sold homes on a regional level. It is likely that the turnover-rate within those municipalities is affected differently by different factors. This research takes this into account.

A way to take into account those municipality differences is by clustering them at Urbanization level. Urbanization is a CBS definition that clusters municipalities based on the number of addresses per km^2 (CBS, 2015a). In densely and thinly populated areas live different types of people: in densely populated areas live younger people, meanwhile thinly populated areas mainly consist of an older age group (see Figure 3). It is likely that this younger age group has overall less accumulated wealth compared to the older age group. As stated by de Vries (2014), this younger age group is much more likely to have a mortgage outstanding higher than the value of their current home.

Given those differences, expectations about the behavior of those groups can be formulated. In this paper, I will investigate the link between the number of existing home sales and house prices for different urbanization levels. Here I assume that subsequent buyers aged 15-45 are likely to move to a larger house. This is because this age group is more likely to grow income and have children. These expectations are in line with research like Banks, Blundell, and Oldfield (2004) and Han (2013). However, as mentioned, this age group is more likely to have a mortgage value higher than their house value (negative equity) compared to the older age groups. This financial restriction limits their options to move out of their current house. All in all, this means that, for this younger age group, an increase/ decrease in house price matters more for their decision to move. As this age group is concentrated around high urban areas, it is likely that high urban areas respond differently to house price shocks than low urban areas. This effect thus runs through negative equity.

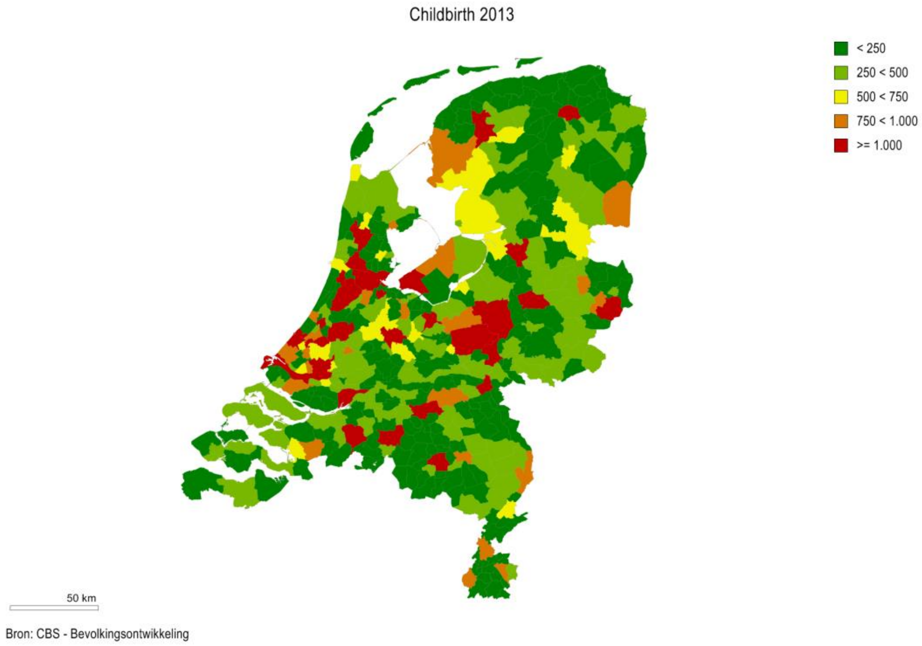
Based on the information above, the 6th hypotheses is:
H6: House price changes affect the number of sold homes more in densely populated areas than in thinly populated areas.

Figure 1: Household Income 2012



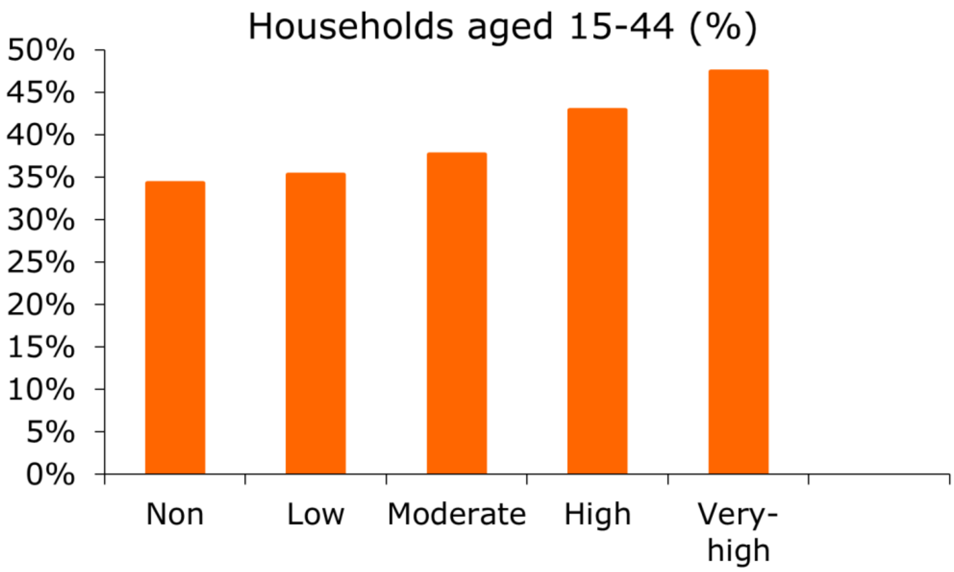
Source: CBS (2015g) Regionale inkomensverdeling, adjustment by Rabobank

Figure 2: Childbirth 2013



Source: CBS-bevolkingsontwikkeling (2015h), adjustment by Rabobank

Figure 3: percentage of households ages 15-44 in different urbanization areas



Source: CBS- inkomensstatistieken (2015i), adjusted by Rabobank

3. Methodology

As previous research indicates, the turnover-rate can be estimated by using its level or first differences. In order to find a comprehensive answer to which determinants explain the number of sold houses in the Netherlands, this research utilizes both methods.

Section 3.1 describes the model in levels. Section 3.2 describes the model in formulated in first differences. Lastly, section 3.3 describes how the stated hypotheses relate to the models.

3.1 Estimating the turnover-rate in levels

As explained in section 2.3, Dutch municipalities differ greatly from each other, both demographically and economically. It is likely that the turnover-rate is affected by those demographic and economic differences between municipalities.

To test whether demographic and economic variables can explain the turnover-rate, it was regressed multiple times from year to year using the first lags of demographic and economic variables. The method used is simple OLS with municipality/urbanization clustered standard errors. The data originally obtained is panel data. However, the data used in this regression is cross sectional, because regressions are run from year to year.

Model 1:

$$(1)\ Turn_{i,t} = \alpha_{i,t} + \beta age1544_{i,t-1} + \gamma \ln(Inc)_{i,t-1} + \delta PPH_{i,t-1} + \varphi \ln(Emp)_{i.t-1} + \varepsilon_{i,t}$$

i indicates clustered municipalities. More information on how municipalities were clustered in the dataset can be found in section 4.5. t is the time in years from 2000-2013. Turnover was defined as the turnover-rate (the number of sold homes divided by the owner occupied housing stock). α is the constant. Age 15-44 was defined as the percentage households with an age between 15-44. Ln(Inc) is the logarithm of average household income. PPH is the average number of people per household. Ln(Emp) is the logarithm of total employment within the COROP area the municipality cluster is in. Lastly, ε is the error term.

3.2 Estimating the turnover-rate in first differences

Another way to estimate the turnover-rate is to regress it in first differences. With first differences it is meant that the value of a variable from last period is deducted from this periods value. This proposed model estimates both *changes* in turnover-rate and house prices over time, making it suited to look more at dynamics of the housing market.

In order to find evidence for the stated hypothesis, the first differences of the turnover-rate and house prices are regressed separately on their first lag and other dependent variables. The model used is a Panel Autoregressive Distributed lag (1) (ADL) model (because it uses the lag of turnover and house prices). The lag of the turnover-rate and house prices are instrumented using their second, third and fourth lag in levels using GMM two stage. The type of data used is panel data, because there are multiple municipalities over time.

Secondly, regressions were run on urbanization level. The general model remains the same as in Model 2, with the difference is that the dataset is restricted to different urbanization types. These regressions were done in order to find whether densely populated municipalities react differently to shocks in independent variables to thinly populated areas.

Model 2:

$$(2)\ \Delta Turn_{i,t} = \Delta\tau_t^1 + \omega_i^1 + \rho^1 \Delta\ln(Pr)_{i,t-1} + \theta^1 Turn_{i,t-1} + \vartheta_{1+j}^1 Int_{t-k} + \sigma_{1+h}^1 \ln(Inc)_{t-l} + \pi^1 \ln(emp) + \in_{i,t}^1$$

$$(3)\ \Delta Ln(Pr)_{i,t} = \Delta\tau_t^2 + \omega_i^2 + \rho^2 \Delta lnPr_{i,t-1} + \theta^2 Turn_{i,t-1} + \vartheta_{1+j}^2 Int_{t-k} + \sigma_{1+h}^2 \ln(Inc)_{t-l} + \pi^1 \ln(emp) + \in_{i,t}^2$$

In this case, i is 331 clustered municipalities as described. t is the time in years from 2000-2013. τ_t are differenced time fixed effects and ω_i are clustered municipality fixed effects. Ln(Pr) was defined as the natural logarithm of house prices. Turn was defined as the turnover-rate. The variable Int corresponds to the average interest rate on new mortgages. Ln(Inc) was defined as the natural logarithm of household income. Ln(emp) expresses the natural logarithm of employment. ε is the error term. For the mortgage interest rate, household income and employment variables, multiple lags can be used as dependent variables. The entire model is stated in first differences.

One has to note that going from model 1 to 2 and 3, all demographic variables disappeared. This has to do with the nature of model 2 and 3. As it is stated in first differences, demographic variables stated in first differences would likely result in insignificant estimates. This is mainly because demographic variables tend to change only little from year to year, meaning that they have very small first differences (as found for example by Dröes and Francke (2016)).

A disadvantage of the proposed model is that it estimates house prices and turnover-rates separately. This means that it does not take account of reverse causality bias (the turnover-rate causes house prices and house prices cause the turnover-rate).

However, this method does take account of several other factors. Firstly, it takes account of multicollinearity within the model. If house prices and turnover-rates estimates are highly auto correlated, then adding more than one lag into the regression will result in high correlations within the model. To minimize this, only the first lag of house prices and turnover were added, and they were instrumented on their second, third and fourth lag using GMM two stage. Secondly, by adding the lags of house prices and the turnover-rate, this method also takes account of possible momentum (as found for example by Dröes and Francke (2016), and Clayton et al. (2010)). Lastly, to take account of data intensity problems the longest possible timeframe on which data is available is used (2000-2013).

A way to take account of reverse causality bias is to estimate both house prices and the turnover-rate together. This can be done using a bivariate Panel Vector Autoregressive model instead of the proposed panel ADL(1) model. In a classic VAR model, all variables used are treated as interdependent and endogenous (Canova & Ciccarelli, 2013). In this specific case, the proposed bivariate PVAR(1) model treats both the number of sold homes and house prices as interdependent and endogenous. Results of this model are not directly shown, but instead can be found in **appendix G.**

3.3 How do the models relate to the stated hypotheses?

- In order to find whether an increase in price will lead to a higher or lower turnover-rate **(hypothesis 1)**, the found coefficient of house prices ρ^1 needs to be significant.
- In order to find a relationship between the turnover-rate and employment **(hypothesis 2)**, a positive significant coefficient needs to be found for π^1.
- In order to find a relationship between the turnover-rate and (one or more) lags of household income **(hypothesis 3)**, a positive significant coefficient needs to be found for σ^1_{1+h} .
- To find evidence for **hypothesis 4**, the coefficient ϑ^1_{1+j} of (one or more lags of) the average mortgage interest rate need to be significantly negative.
- Momentum is found for both house prices and the turnover-rate **(hypothesis 5)** if the coefficient $\rho^2\theta^1$ is significantly positive.
- Price changes affect the turnover-rate more in densely populated areas than in thinly populated areas **(hypothesis 6)** if the effect of prices on the turnover-rate (ρ^1) is (significantly) higher for municipality regressions with high urbanization compared to low urbanization.

4. Data Description, descriptive statistics and testing

In order to find evidence for the stated hypotheses, data needs to be collected. The preferred timespan is 2000-2013, because this is the longest period on which data is available for all variables (except housing stock). In all cases, yearly data is used on a municipality level. This is done, because quarterly level data is mostly unavailable on municipalities. All variables are gathered on a nominal level. This is because research like Dröes and Francke (2016) find no evidence that inflation plays a big role in explaining house prices and the number of sold homes. The obtained dataset is panel data: it contains of multiple variables over time.

The structure of Section 4 is as follows. Section 4.1 elaborates on the only national variable used. This is a variable equal to all municipalities. Section 4.2 and 4.3 describe the municipality economic and municipality demographic data respectively. Section 4.4 describes how urbanization can be used to compare small towns to big cities. Section 4.5 explains how municipality data will be merged in order to create a more reliable estimation of house prices and the number of house sales. 4.6 Shows descriptive statistics. 4.7 conducts stationarity tests. Lastly, 4.8 discusses correlation statistics. An extensive outlier analysis can be found in **appendix A. From now on, all data is shown without outliers.**

4.1 National Variable

The obtained national variable is the mortgage interest rate on new mortgages. The mortgage interest rate is calculated by grouping interest rates on newly issued mortgages into four groups: variable and fixed interest rates with less than 1 year to maturity, fixed interest rate 2-5 years to maturity, fixed 5-10 years to maturity and fixed more than 10 years to maturity. Given the volume of these mortgages, the groups are averaged to create the average interest rate on newly issued mortgages.

Variable sources and used timespan can be found in Table 2.

4.2 Municipality economic data

Municipality economic data obtained are the number of existing home sales, house prices, housing stock, employment, and household income. The number of existing home sales and housing stock is used to create the turnover-rate. The turnover-rate makes municipalities more comparable as some low housing stock municipalities have naturally less transactions.

The number of existing home sales and average house prices were obtained from in house Land Registry data at the Rabobank. Both house prices and the number of excising home sales are divided into 5 house types: apartments, mid-terrace, end of terrace, detached-, semidetached-, and unknown. For some years in certain municipalities no sales occurred for one or more house types. For those years, the number of transactions is set to 0. House prices in that municipality are then set to missing for that house type in that year.

Housing stock has been obtained from CBS Statline. Housing stock has been defined as the number of owner occupied houses per municipality (excluding rental homes). As the data is only available in the municipality 2012 division, it had to be reconstructed to fit the CBS 2015 municipality definition used in all other variables. A detailed explanation of how the data was reconstructed can be found in **Appendix B**. Data on housing stock is available from 2006-2012. This means that it does not fit the preferred 2000-2013 timeframe. In order to fit the data to the 2000-2013 timeframe, housing stock was linearly extrapolated. This was done both backward to 2000 and forward to 2013. Unfortunately, some municipalities experienced growth of such magnitude that extrapolation resulted in a negative housing stock for at least one year. In addition to this, one municipality had too little data points (1) to conduct extrapolation. For this reason, the municipalities Hollands Kroon, Binnenmaas, Medemblik, Nieuwkoop, Roerdalen, and Roermond were deleted from the dataset. This resulted in a total of 387 municipalities.

Employment was obtained from Landelijk Informatiesysteem van Arbeidsplaatsen (LISA). Employment was defined as the amount of jobs available within a certain municipality. This means that it measures the amount of jobs available, not the amount of employed people living in a municipality. Ultimately, this could result in a distorted picture as the decision to move is based on whether a household is employed, not whether there are many jobs available in the municipality. To reduce this problem, the sum of employment was calculated within a COROP area. As it is most likely that people live and work within one COROP area, this figure is expected to better reflect the amount of people employed within a municipality.

Lastly, Income was obtained from CBS Statline. It was defined as the average household income within a municipality.

4.3 Municipality demographic data

Through CBS (2015b), data on 393 different municipalities were collected. Those 393 municipalities are in line with CBS 2015 grouping (CBS, 2015b). Obtained demographic data on municipalities are household age, the amount of people per household and childbirth.

Household age was obtained from Statistics CBS Statline. Household age was (in most cases) defined as the age of the adult male living in the house. Given that the male is most likely to be oldest person within the household, this results in a higher household age than if one would take the average age of all people within a household. Household age was divided into 10 year groups, ranging from 15-24 to 95+. Percentages were then calculated for each municipality.

Other variables gathered from CBS Statline are the number of people per household and childbirth. The number of people per household was defined as the average number of people within one household in a certain municipality. Childbirth was defined as the amount of children born in a municipality.

Table 2: Variable sources and timespan

	Timespan	**Source**
National Data		
Mortgage Interest Rate	2000-2013	The Nederlandsche Bank
Municipality economic		
Existing Number of Home Sales	2000-2013	In house land registry data at the Rabobank
Transaction Price	2000-2013	In house land registry data at the Rabobank
Housing Stock	2006-2012	CBS Statline
Employment	2000-2013	'Landelijk Informatiesysteem van Arbeidsplaatsen (LISA)'
Household Income	2000-2013	CBS Statline
Municipality demographic		
Municipalities	2015 definition	CBS Statline
Urbanization		CBS Statline
Household age	2000-2013	CBS Statline
Amount of people per household	2000-2013	CBS Statline
Childbirth	2000-2013	CBS Statline

Notes: For all municipalities, the 2015 CBS grouping is used (CBS, 2015b).

4.4 Urbanization

In order to discover whether house price changes affect the number of sold homes more in densely populated areas than in thinly populated areas **(hypothesis 6)**, big cities need to be compared to small towns. To make a distinction between those big cities and small towns, the variable urbanization was chosen. Urbanization is a CBS definition that clusters municipalities based on the number of addresses per km^2 (CBS, 2015a). Urbanization ranges from non-, low-, moderate, and high to very high urban (CBS, 2015a). High urban areas contain a lot of homes close to each other. Meanwhile, low urban areas contain few houses close to each other. Given this fact, urbanization is a proxy of population density. The proportions of urbanization types throughout the dataset can be found in Table 3. As can be seen in Table 3, most municipalities are low-urban (37.22%), meanwhile the least are very-high urban (2.84%).

Table 3: 5 types of urbanization

Urbanization	Percentage before merging	Percentage after merging
Non-urban	22.44%	18.87%
Low-urban	37.22%	28.30%
Moderate-urban	22.73%	28.30%
High-urban	14.77%	18.87%
Very High-urban	2.84%	5.66%
Total	**100%**	**100%**

4.5 Merging

Working with data on a municipality level resulted in several problems. As used the data is on municipality level, the absolute number of transactions can be low when municipalities are small. This results in two problems. Firstly, the houses types (mid-terrace, apartment, etc.) being sold from one year to the next can differ substantially. Secondly, the quality of those homes can change from year to year too. Because of the low number of home sales, the differences between house quality and house types over the years don't even out as they would if the number of sold homes were high. As a result, the average price can be volatile from year to year in small municipalities. This results in an unreliable price estimates.

At least for some municipalities, the number of home sales and the average house prices are too little to do significant statistical inference. In this particular dataset, the number of home sales for any municipality can be as small as 4, but as large as 8562 (with a median of 289). An (extreme) example can be shown for Schiermonnikoog (Figure 4). The tradeoff here is obvious: to obtain more trustworthy price estimates, one has to give up observations. In other words: municipalities should be merged for appropriate econometric techniques.

Merging municipalities can be done in several ways. Firstly, one could merge municipalities on their urbanization level. Urbanization is a CBS definition that clusters municipalities based on the number of addresses per km^2 (CBS, 2015a). This means that merging municipalities using urbanization can be quite helpful when we want to compare big cities to small towns. A downside to this is however that urbanization only has 5 types. This means that merging municipalities given their urbanization levels will ultimately result in a total of 75 observations (15 observations per year per type of urbanization for each variable).

A second option could be using COROP areas. The Netherlands has 40 COROP areas (CBS, 2015f). This ultimately will result in a total of 560 observations (40 COROP areas times and 14 years), and will enable us to construct a more robust model compared to using urbanization. However, using COROP areas come with another downside. Namely, COROP areas are designed to have a core area with a periphery: an area in which the population works and lives. As a result, COROP areas include both urban and non-urban areas. If we were going to be estimating a national model, using COROP areas would be adequate. However, if our goal is explain differences between urban areas and country sides, using COROP areas will destroy our ability to do so. Therefore, using COROP areas alone are not a viable option.

A more feasible option would be to combine the definitions of COROP areas and urbanization to merge municipalities. Using this method, all municipalities within a certain COROP area were merged, provided they have the same type of urbanization. In theory, this could result in 200 clusters (5 types of urbanization times 40 COROP areas) with a maximum of 2800 observations (200 clusters times 14 years). However, not every COROP area contains municipalities of all 5 urbanization types. The actual amount of clusters obtained in this dataset is 131 (106 after deleting outliers).

Merging on COROP and urbanization provides us with some good results. As figure 5 shows, clustering Schiermonnikoog together with other non-urban areas in the same COROP area resulted in significantly more observations and less volatile transactions. As can be seen from Table 3, the amount of areas with a low-urban profile decreased after merging, meanwhile high-urban and very high-urban areas increased. This indicates that those problem areas are clustered less often than low-urban areas.

All combined municipalities can be found in **Appendix C**.

Figure 4: Number of home sales and average house prices in Schiermonnikoog before merging

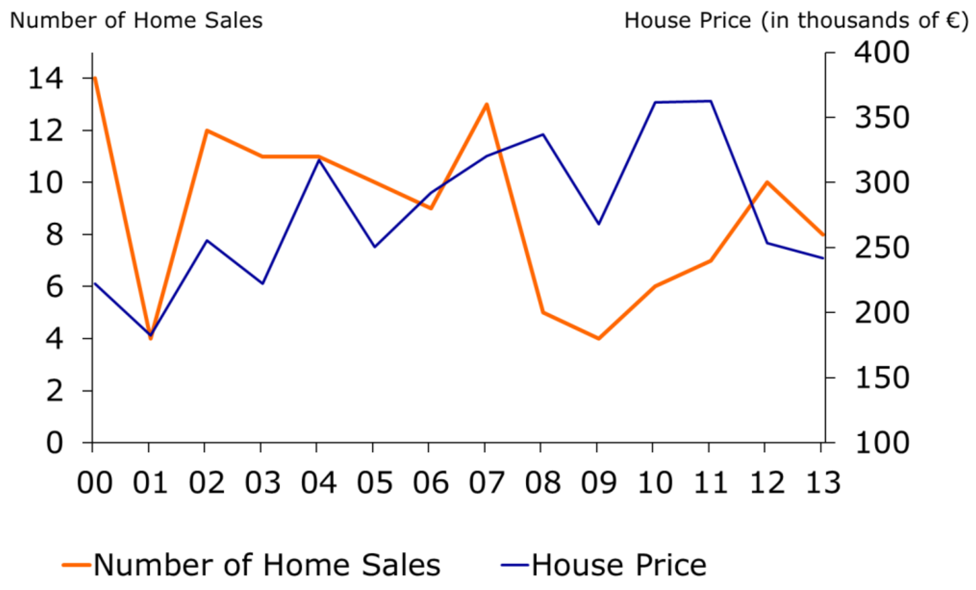
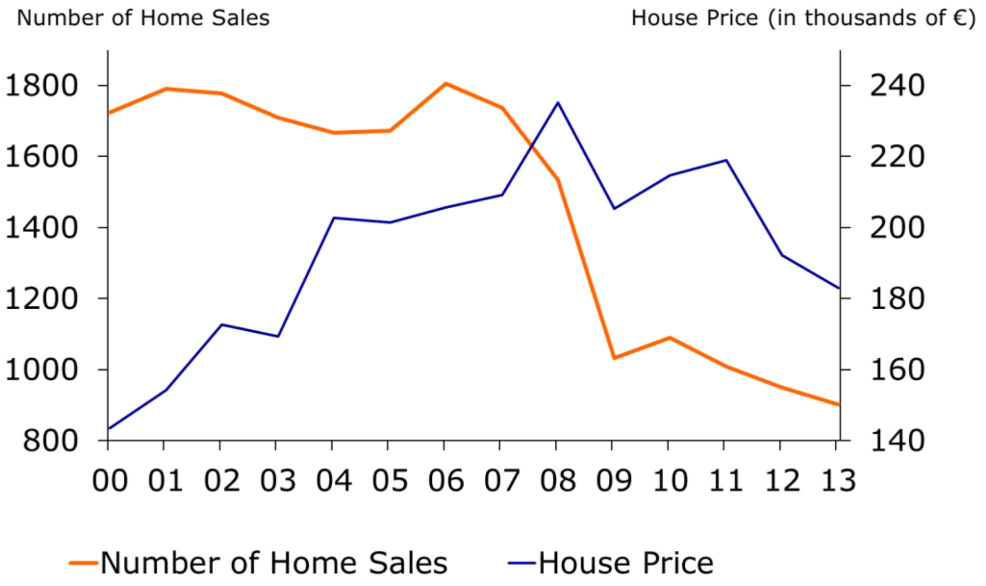


Figure 5: Number of home sales and average house prices in the cluster of Schiermonnikoog after merging



4.6 Descriptive Statistics

This section covers descriptive statistics. It was divided into three sections. 4.6.1 covers the national variable. Section 4.6.2 and 4.6.3 covers municipality economic and municipality demographic variables respectively. Descriptive statistics of the variables in levels can be found in Table 4. From variables figures were made. These were reported in Figure 6-10. In addition to this, descriptive statistics of variables in first differences were made. Since these are harder to interpret, they are not directly discussed. The descriptive statistics of variables stated in first differences can be found in **Appendix D.**

4.6.1 Descriptive statistics: national variable

As shown by Figure 6 and Table 4, the mortgage interest rate fluctuated around 3.70% to 5.88%. The mortgage interest rate mainly decreased throughout the dataset.

4.6.2 Descriptive statistics: municipality economic variables

Used Municipality economic variables are the number of home sales, house prices, turnover, employment and average household income.

As shown by Figure 7, house prices rose significantly between 2000-2008, but decreased again after. This house price drop can be mainly explained by the economic situation during financial crisis and new tighter credit conditions on the mortgage market.

The turnover-rate shows a slightly different trend to house prices. As can be seen from Figure 8, between 2000-2004, the turnover-rate was stable. But, this dropped significantly between 2007-2013 for all urbanization levels. On average, the turnover-rate is the highest in very-high urban areas. The lowest turnover-rate is in non-urban areas. Surprising is the difference between median and mean values of house sales per house type. As this research does not directly focusses on those differences, descriptive statistics and a discussion on this topic can be found in **appendix E**.

Figure 9 shows the average employment in a single COROP area. As can be seen from the graph, employment has been mainly increasing from 2000, but has decreased from 2009 onward. Employment can be very different from one municipality to another. For example, employment in 2015 was 450 for Rozendaal. Meanwhile, it was 586 thousand for Amsterdam.

Over time, income has been generally increasing (Figure 10). Decreases in income are very rare, with only small income drops for some urbanization types in some years (2002-2003 and 2008-2011). The highest average income is not earned in the very-high urban areas, but rather in the non-, moderate- and high-urban areas. The lowest income is earned in very high- and low- urban areas. In 2013, the lowest average yearly household income (28.3 thousand Euros) was earned in Heerlen. Meanwhile, the highest average yearly household income (58.3 thousand Euros) was earned in Rozendaal in the same year.

4.6.3 Descriptive statistics: municipality demographic variables

The amount of people per household is mostly constant in The Netherlands, with some outliers: the lowest average amount of people per household in 2015 is 1.66 in Groningen, meanwhile Urk has on average the largest amount of people per household (3.37). On average, most births are in Amsterdam and Rotterdam. The least amount of children being born are in Rozendaal. The oldest people (aged 75-95) generally live in Heemstede, meanwhile the youngest age group (aged 15-44) prefers to live in Utrecht and Groningen.

Figure 6: Mortgage Interest Rate (2000-2013)

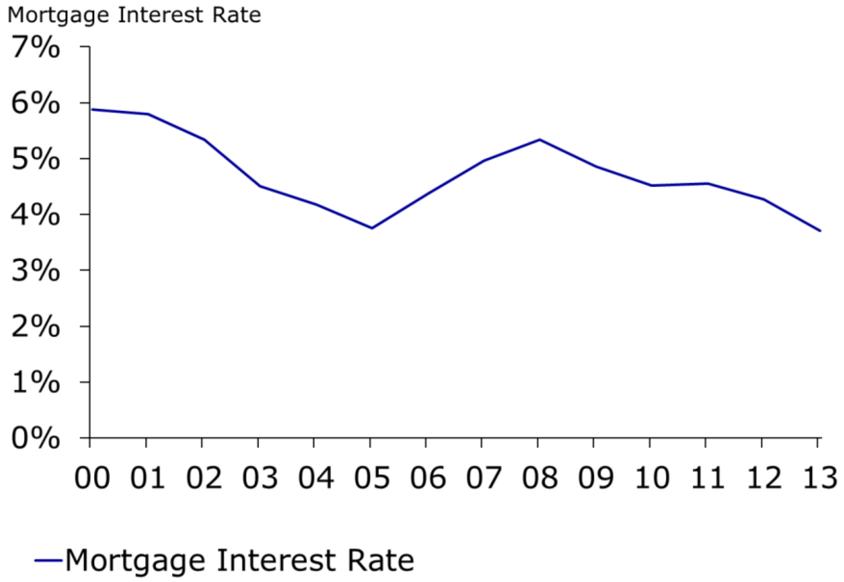


Figure 7: Average house prices (2000-2013)

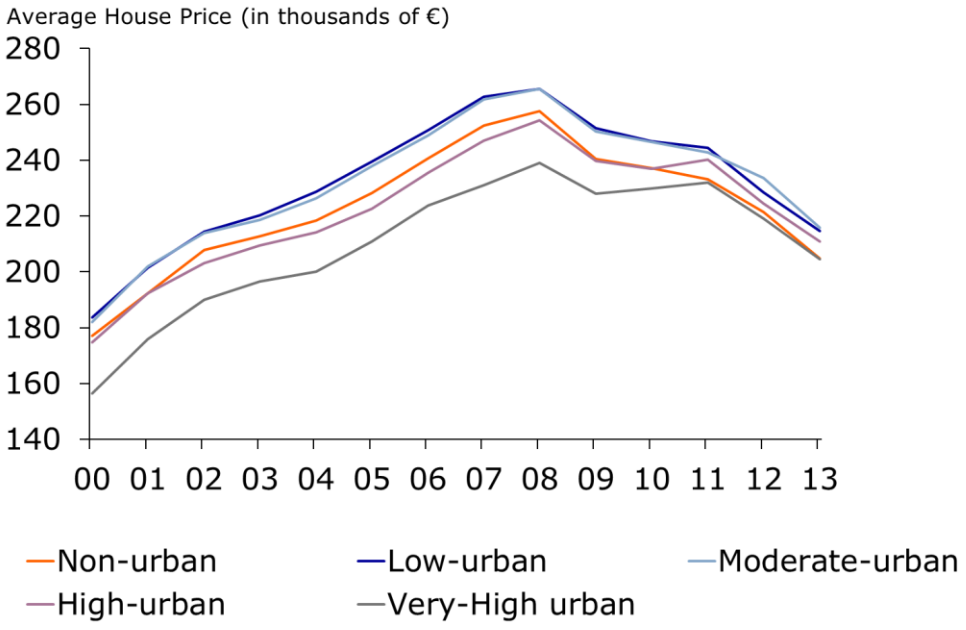


Figure 8: Average Turnover-rate (2000-2013)

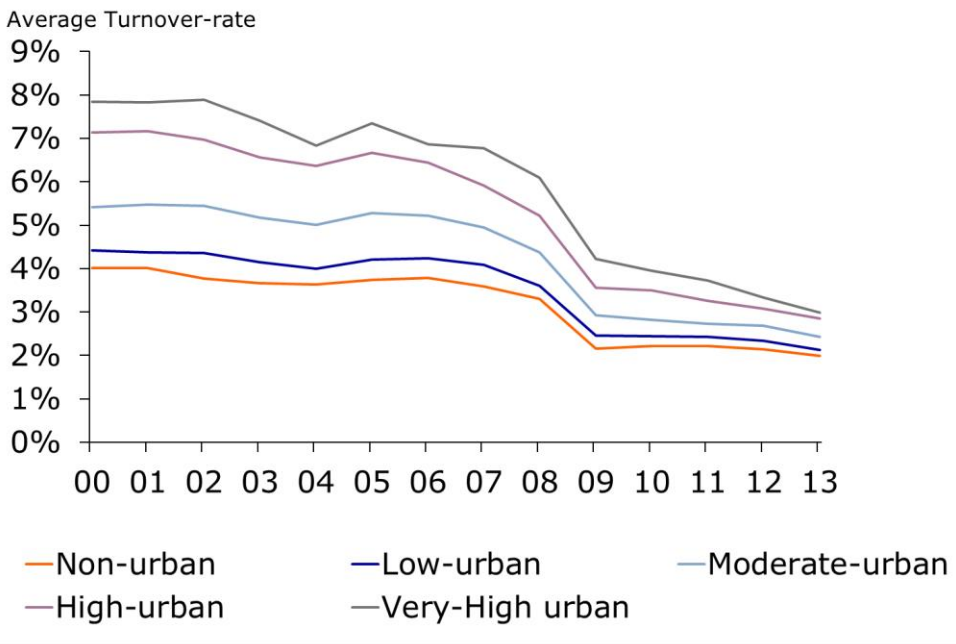


Figure 9: Employment within a COROP area (2000-2013)

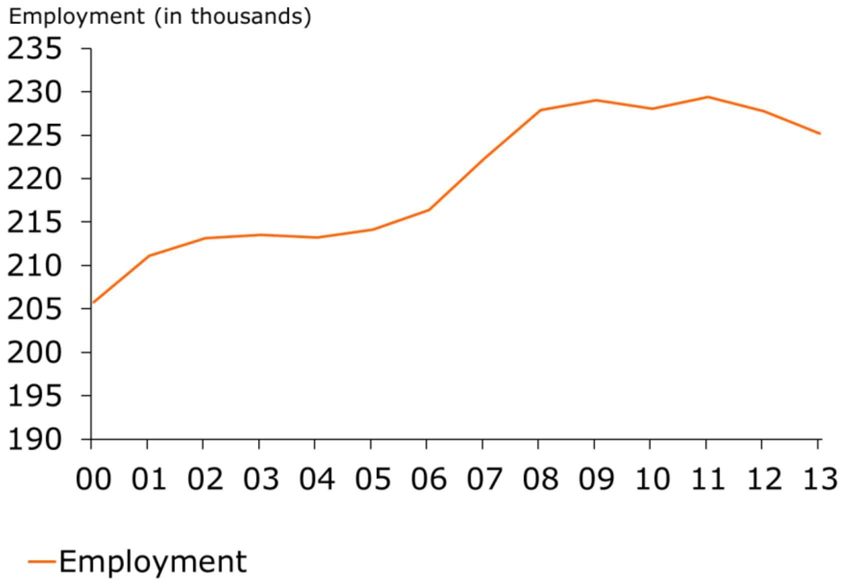


Figure 10: Average Household Income (2000-2013)

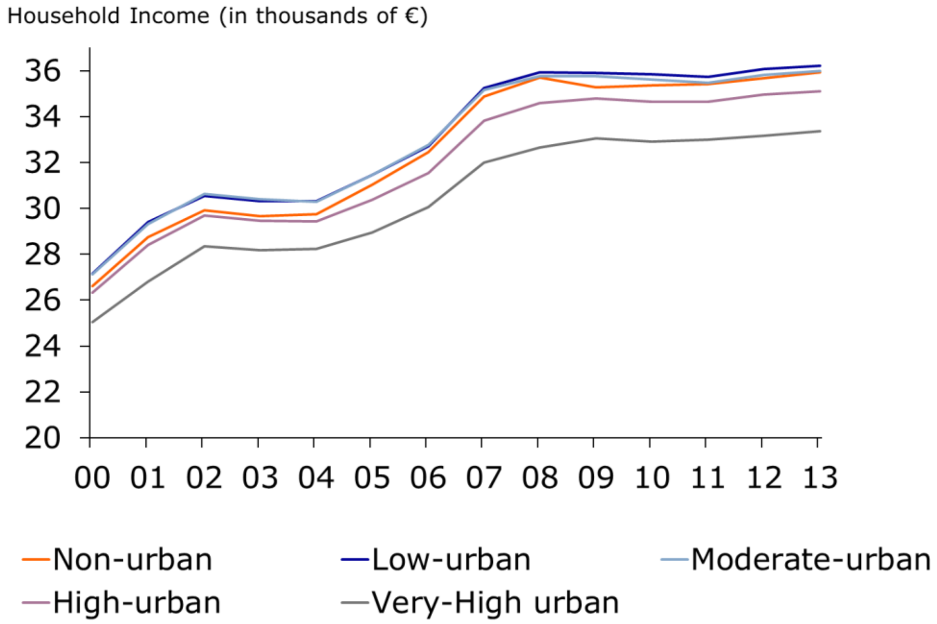


Table 4: Descriptive Statistics of variables stated in levels (Annual data 2000-2013)

	Mean	**Median**	**N**	**Std. Dev**	**Min**	**Max**
National Data						
Mortgage Interest Rate (in %)	4.72	4.54	14	0.68	3.70	5.88
Municipality economic						
Number of Existing Home Sales	1281	986	1484	1290	42	12450
Transaction Price (Thousand €)	227	221	1484	52.40	92.88	495
Housing Stock (Thousand)	29.84	25.65	1484	22.51	2906	167
Turnover-Rate (in %)	4.15	3.91	1484	1.60	1.38	9.88
Employment (Thousand)	220	163	1484	176	181	843
Household Income (Thousand €)	32.59	32.35	1484	4.24	22.60	48.80
Municipality demographic						
Household Age (per group in %)						
- 15-24	3.48	2.50	1484	2.72	1.00	19.00
- 25-34	14.12	16.35	1484	3.44	6.00	26.00
- 35-44	20.30	20.41	1484	2.09	14.00	30.00
- 45-54	20.48	20.57	1484	1.63	15.00	25.00
- 55-64	17.76	18.00	1484	2.32	9.00	22.50
- 65-74	13.03	13.00	1484	1.98	6.00	19.00
- 75-84	8.71	8.75	1484	1.41	4.00	14.00
- 85-and up	2.42	2.28	1484	0.71	1.00	5.00
People per Household	2.42	2.42	1484	0.19	1.79	2.95
Childbirth	1411	1147	1484	1232	93	9748

Notes: This Table shows descriptive statistics for the variables used in this research. They were divided into three categories: National-, municipality demographic-, and municipality economic data. All data shown is from 2000-2013 and is without outliers.

4.7 Stationarity tests

Given the considerable chance that used variables are in fact non-stationary, panel unit root tests were done. The test chosen is a Fisher-type test based on the Phillips-Perron test. This type of test is suited for the panel data used in this study. The test includes the first lag and time trends. The formal notation of this test is stated in equation 3. Results are shown in Table 5. In this case, the null hypothesis is that all panels contain unit roots (are non-stationary). The H1 hypothesis is that at least one panel is stationary.

$$[3]\ y_{i,t=}\alpha + \rho\gamma_{t-1} + \delta_t + u_{i,t}$$

Overall, the test gives overwhelming evidence that all level variables contain unit roots. In all cases the p-value tends to 1. This result however relies heavily on whether a time trend is included.

The same Fisher-type test is also done on variables stated in first differences. The test is conducted in exactly the same way as for the variables stated in levels. These tests indicate that almost all variables are stationary in differences. An exception to this is the mortgage interest rate, which is still non-stationary, according to the inverse Chi-squared test. A reason for this could be that mortgage interest rates have been decreasing since 2007 in this dataset. However, even though technically speaking the mortgage interest rate has no upper nor lower bound, in reality this is not the case. The mortgage interest rate is unlikely to go much lower than 0% (for a long time), and cannot increase infinitely. Therefore, it is likely that the mortgage interest rate is actually in fact stationary, even though the tests do not point this out. Given the overwhelming evidence that (almost) all variables are stationary in differences, the (differenced) turnover-rate regression will be estimated entirely in first differences.

Table 5: Stationarity tests

	Levels		Differences	
Variable	Inverse Chi-sq.	p-value	Inverse Chi-sq.	p-value
Mortgage Interest Rate	63.00	1.00	84.58	1.00
Turnover-Rate	126.51	1.00	677.09	0.00
Ln(Transaction Price)	12.35	1.00	1104.74	0.00
Employment	70.35	1.00	286.16	0.00
Household Income	89.42	1.00	304.97	0.00

Notes: This table shows stationarity tests for all variables used in model 2 and 3. The test used is a Fisher-type test based on the Phillips Perron test. Timeframe used is 2000-2013. Data is shown without outliers.

4.8 Correlations

Table 6 shows correlations between variables stated in levels. Both municipality demographic and municipality economic variables are included in this table. Most variables are moderately correlated with each other, but nearly not high enough to expect imperfect multicollinearity within the model. As can be seen from Table 6, the correlation with its lag is extremely high (0.94). This indicates that the absolute value of the turnover-rate actually does not change that much. Most interesting however are the other correlations with the turnover-rate, as this will be de dependent variable in the regression. The turnover-rate is significantly correlated with the household age 15-44 (0.63). This could indicate that people aged 15-44 move more often. Income is negatively correlated (-0.59) with the turnover-rate. This could indicate that people with a higher income tend to move less often. This is against expectations, as we would expect an increase in income to have a positive effect on the turnover-rate **(hypothesis 3).** The correlation coefficient between childbirth and turnover-rate is significantly positive (0.33). This could mean that people tend to move more often after their child was born instead of building an addition to their house to ensure extra space. The correlation between the turnover-rate and the amount of people within a household is significantly negative (-0.27). This indicates that the more people live in one house, the less often they tend to move. Lastly, the correlation coefficient between the turnover-rate and employment is insignificant (-0.02).

Table 7 shows correlations for variables stated in first differences. Most variables are correlated with each other, but again not high enough to suspect imperfect multicollinearity within the model. The correlation coefficient between house prices and the lag of house prices and turnover-rate is significantly positive and not significantly positive respectively. As previous literature predicts a (positive perhaps negative) relationship, this is in line with expectations **(hypothesis 1)**. However, as the model will include the lag of house prices, the insignificant correlation coefficient might result in insignificant estimates within the model. The turnover-rate is significantly negatively correlated with its lag. This is against **hypothesis 5**, which states that the turnover-rate might be subject to momentum**.** This is not the case when an increase in the turnover-rate results in a decrease in the turnover-rate in the next year. The correlation coefficients between employment and the turnover-rate is positive and not statistically significant **(hypothesis 2).** The insignificance of this correlation coefficient is not in line with expectations, as former research suggest a positive significant relationship. The correlation coefficients of household income has the expected positive sign, indicating that an increase in household income often goes together with an increase in turnover-rate **(hypothesis 3).** The first and second lag of household income are however negatively correlated with the turnover-rate. This is against expectations. In line with expectations, the coefficient of mortgage interest rate and turnover-rate is significantly negative **(hypothesis 4).** All in all, correlation show how dynamic the data actually is. Some variables are significantly correlated with the turnover-rate, but not when using their lags. Other variables are significantly correlated with the turnover-rate when using lags but switch signs.

Table 6: Correlations between variables (stated in levels)

	Turnover-Rate	Turnover-Rate$_{t-1}$	Household Age15-44$_{t-1}$	Ln(Income)$_{t-1}$	Ln(Childbirth)$_{t-1}$	People per Household$_{t-1}$	Ln(Employment)$_{t-1}$
Turnover-Rate	1						
Turnover-Rate$_{t-1}$	0.95***	1					
Household Age 15-44$_{t-1}$	0.63***	0.64***	1				
Ln(Income)$_{t-1}$	-0.59***	-0.53***	-0.47***	1			
Ln(Childbirth)$_{t-1}$	0.33***	0.34***	0.42***	-0.11***	1		
People per Household$_{t-1}$	-0.27***	-0.31***	-0.25***	0.19***	-0.17***	1	
Ln(Employment)$_{t-1}$	-0.02	0.03***	0.11***	0.30***	0.53***	0.09***	1

Notes: This Table shows correlations of variables used in model 1. Timeframe used is 200-2013. Data is shown without outliers.

** Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level*

Table 7: Correlations between variables (stated in first differences)

	∆Interest Rate $_{t-1}$	∆Turnover-Rate	∆Turnover-Rate$_{t-1}$	∆Ln(Price)	∆Ln(Price)$_{t-1}$	∆Ln(Income)	∆Ln(Income)$_{t-1}$	∆Ln(Income)$_{t-2}$	∆Ln(Employment)
∆ Interest Rate$_{t-1}$	1								
∆Turnover-Rate	-0.39***	1							
∆Turnover-Rate$_{t-1}$	0.10***	-0.05*	1						
∆Ln(Transaction Price)	-0.09***	0.32***	0.19***	1					
∆Ln(Transaction Price)$_{t-1}$	0.18***	0.03	0.32***	0.306***	1				
∆Ln(Income)	0.42***	0.24***	0.30***	0.54***	0.38***	1			
∆Ln(Income)$_{t-1}$	0.48***	-0.08***	0.24***	0.35***	0.53***	0.37***	1		
∆Ln(Income)$_{t-2}$	0.21***	-0.45***	-0.09***	0.05*	0.32***	-0.22***	0.36***	1	
∆Ln(Employment)	0.45***	0.008	0.19***	0.40***	0.35***	0.48***	0.50***	0.20***	1

Notes: This Table shows correlations of variables used in model 2 and 3. Timeframe used is 200-2013. Data is shown without outliers. ∆ Indicates that the first difference has been taken. Indicates a significance at 10% level ** Indicates a significance at 5% level *** Indicates a significance at 1% level.*

5. Empirical Results

This section elaborates on the empirical results found in this research. Firstly, section 5.1 explains the results of the turnover-rate model in levels (model 1). Then, for those results it's discussed whether they are in line with previous studies. To give a little bit more depth to the research, it is investigated in section 5.1.1 whether coefficients are stable over time. Section 5.1.2 then investigates how well the turnover-rate in levels predicts. Section 5.2 elaborates on the turnover-rate and house price model stated in first differences (model 2 and 3). It was then elaborated on whether results are in line with other studies and hypotheses. Section 5.2.1 states the turnover-rate and house price model in differences for different urbanization levels.

5.1 Turnover-rate model in levels

Table 8 shows one of the 13 regressions done to estimate the turnover-rate in levels. Details of said regressions can be found in **appendix F**. The overall R-squared of the regressions is between 0.553-0.728. In addition, F-tests were done to check whether the variables together are significant. F-values in all years are significantly positive using a 99% confidence interval. This indicates that all variables together statistically significantly explain variations in the turnover-rate. Looking at the interpretation of regression 5, one should notice that the turnover-rate itself is small. In this dataset, the turnover-rate varies between 1.4%-10% for municipality clusters. Therefore, the found coefficients in the regressions will be small as well.

The coefficient of Age1544 (percentage population between 15-44) has a positive sign and is statistically significant using a 99% confidence interval for all regressions. If in 2005 the number of people aged 15-44 within a municipality cluster is on average 10% higher, then the turnover-rate is estimated to be 0.0095 higher. This indicates that people aged 15-44 move significantly more often than others. An explanation to this phenomenon could lay in the fact that this age group is a little bit "more dynamic" can others. Namely, this age group is more likely to move out of their parents' house, find a life partner, obtain a job, and have kids. Only the latter is corrected for by taking into account the number of children being born. Meanwhile, the older age group already has gone through these changes, and thus likely has a lesser need to move house. This tendency to move more often will result in higher turnover-rates for the age group 15-44. These findings are in line with expectations like Dröes and Francke (2016), who hypothesize the positive age 15-44 effect but did not find evidence. They ascribe this lack of found evidence to their model being stated in first differences (not levels).

Table 8: Turnover-rate model in levels

	(5)
	Turnover-Rate
Date	2005
$Age1544_{t-1}$	0.095***
	(0.017)
$Ln(Income)_{t-1}$	0.019
	(0.013)
$Ln(ChildBirth)_{t-1}$	0.004***
	(0.001)
People per $Household_{t-1}$	-0.039***
	(0.005)
$Ln(Employment)_{t-1}$	-0.002
	(0.001)
Constant	-0.090
	(0.117)
N	106
adj. R^2	0.693
F	28.28***
RMSE	0.008

*Notes: Regressions used are simple OLS with clustered standard errors. Regressions are on the entire sample (2000-2013) and excludes outliers. Ln indicates that the natural logarithm is taken * Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level.*

The coefficient of household income is positive. This variable is not statistically significant in 2005 using a 90% confidence interval. It however does become significant in the year 2004 and from 2006 onwards. The positive coefficient could indicate that people with a higher income tend to move slightly more often. This could potentially mean that income has started to play a bigger role when obtaining a mortgage, since generally speaking mortgages are more easily obtained for households with higher income. Thus, people with high household income move more often simply because they can. These results are in line with previous research like Dröes and Francke (2016), Clayton et al. (2010) and Fisher et al. (2004). However, one has to notice that Clayton et al. (2010) and Dröes and Francke (2016) investigated the turnover-rate in using, and Fisher et al. (2004) investigated commercial properties using a probit model. Therefore this research might not be directly comparable with this regression.

The coefficient of childbirth is significantly positive using a 99% confidence interval throughout all regressions. The coefficient is 0.004 in 2005, indicating that the turnover rate is expected to be on average 0.0008 higher for municipality clusters that had on average

20% more children born the year before. This means that it is likely that households move more often after their family is expanded. A reason for this could be that with family increases, a larger home is needed. Instead of expanding the home, a significant number of families decide to just move house.

Adding the amount of people per household to the regression results in a significantly negative coefficient. If within a municipality cluster on average lived 1 person extra within a household, then the turnover-rate is expected to be 0.039 lower in 2005. This is quite a significant drop given that the average turnover throughout the Netherlands is only 4.2%. These results could mean that large families tend to move less often. An intuitive explanation for this could be that it is simply more difficult to move with a larger family. In addition, since there are less homes suitable for large families, they also simply have less choice in houses. Another reason could lay in household income. If a family with average household income includes more kids, then a larger proportion of this income will be put to taking care of children. Hence, there will be less income left for moving. Thus, it could also be the case that these families move less often, because their income does not allow for this. As this research does not directly investigate how much income families with many children earn, the last mentioned effect cannot be proven. This thus needs to be taken as a suggestion needing more investigation.

The employment within a COROP area provides a negative coefficient. Employment is only found to be statistically significant using a 90% confidence interval in the year 2010. The overall effect is of employment is relatively small, and evidence about its significance over time points to the employment rate not being significant. Strictly speaking this is against research like Clayton et al. (2008) who find a significant positive effect of employment on the turnover-rate. However, because employment is added to the regression in levels, and not first differences, it might not be directly comparable.

5.1.1 Turnover-rate model in levels- How consistent are the coefficients over time?

Even though most regression coefficients were found to be statically significant, they will be of little use predicting the turnover-rate when they are subject to considerable change from year to year. Therefore, Figure 11 and 12 report the coefficients of the turnover-rate equation in levels throughout the years. Figure 11 shows the coefficients of child birth and employment. Figure 12 shows the age group 15 to 44, household income and the average amount of people per household.

Figure 11: Turnover-rate model coefficients of Childbirth and Employment over time (2000-2013)

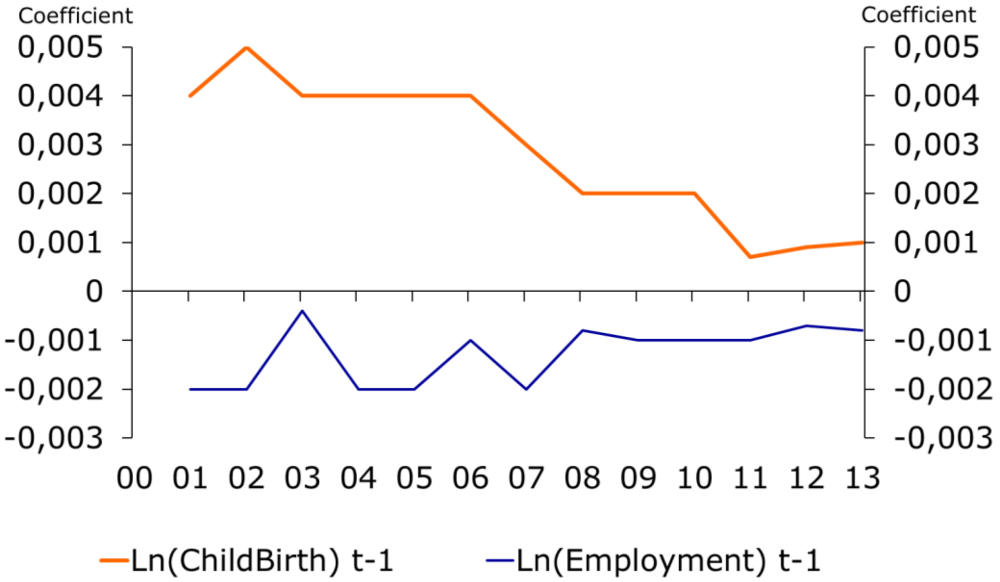
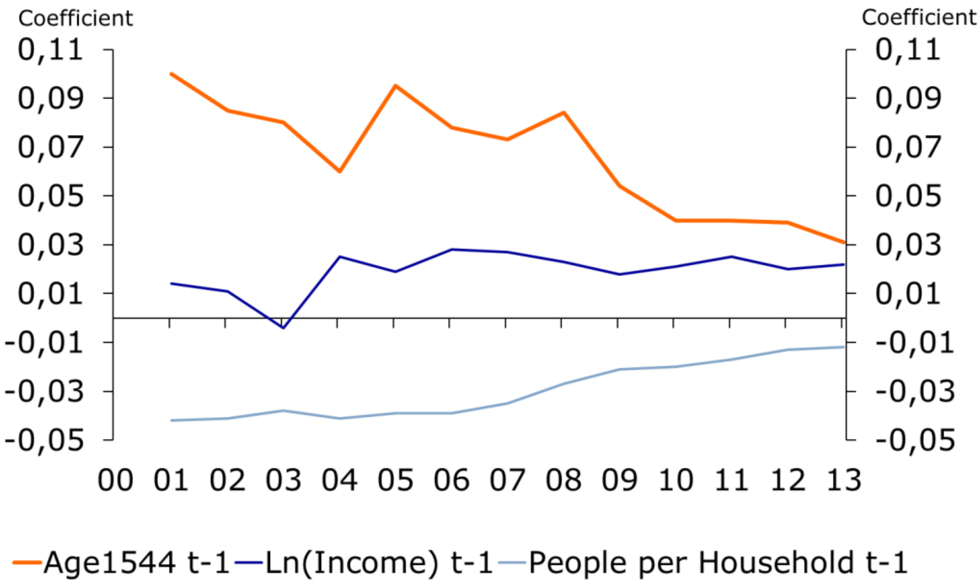


Figure 12: Turnover-rate model coefficients of the percentage age 15-44, income and the amount of people per household over time (2000-2013)



As can be seen from figure 11 and 12, most coefficients change over time. For example, the coefficient of the average amount of people per household increases throughout time, meanwhile the coefficient of child birth seems to have a decreasing trend. Coefficients themselves however are not very volatile. However, coefficients themselves are quite stable throughout time. Even when there is an up- or downward trend, the coefficient from last year will not be very different from this year.

5.1.2 Turnover-rate model in levels- How well does the model predict?

Now we know that demographic and economic variables are likely related to the turnover-rate stated in levels, a question that could be asked is: how well can this model predict next year's turnover-rate? To answer this question, for each year a prediction is made for the next year (the 2001 regression using dependent variables from the year 2000 predicts 2002, the 2002 estimation using dependent variables from the year 2001 predicts 2003 etc.). To check whether the model predicts well, Figure 13 plots the models predicted turnover-rate together with the actual turnover-rate and the lag of the turnover-rate. Overall, the model predicts the turnover-rate quite well. Before and after 2009 the model is most close to the actual turnover-rate. In cases where the turnover-rate decreased, the model seems to slightly over-predict the turnover-rate. Meanwhile, when the turnover-rate increased, the model seems to slightly understate the turnover-rate. Around 2009 the model starts to deviate from the actual turnover-rate. This is to be expected. Given the large dip in the turnover-rate around that time, it is unlikely that demographic variables (which remain quite constant over time) can explain such differences.

However, most comparable to the models predicted turnover-rate is the lag of the turnover-rate. Taking a step back will explain why this is the case. This model uses solely municipality demographic and economic data. Especially municipality demographic data does not change much over time. Only employment and income are (slightly) sensitive to economic cycles. If you then for example estimate the model in 2005, each regression will explain the turnover-rate in 2005 using independent variables from 2004. If you then try to make a prediction of 2006, independent variables from the year 2005 would be used. However, if said independent variables are not subject to big changes from year to year, then you will get almost exact the same turnover-rate estimation using independent variables in the year 2004 and 2005. Hence, you are actually just estimating the lag of the turnover-rate.

Now we know the dynamics of the turnover-rate model stated in levels, one could ask whether the model works well for all municipality clusters. In order to give an answer to this question, the forecast error of the model is computed:

$$Forecast\ error\ =\ ((models\ estimated\ turnover - actual\ turnover)/\ actual\ turnover).$$

This means that if the forecast error is 0.1, then the model is 10% off from the actual turnover-rate. For each year, the minimum, p5, median, p95, maximum and standard deviation of the forecast error is computed. Results can be found in Figure 14.

Figure 13: Predicted turnover-rate compared to the actual turnover-rate and the lag of the turnover-rate

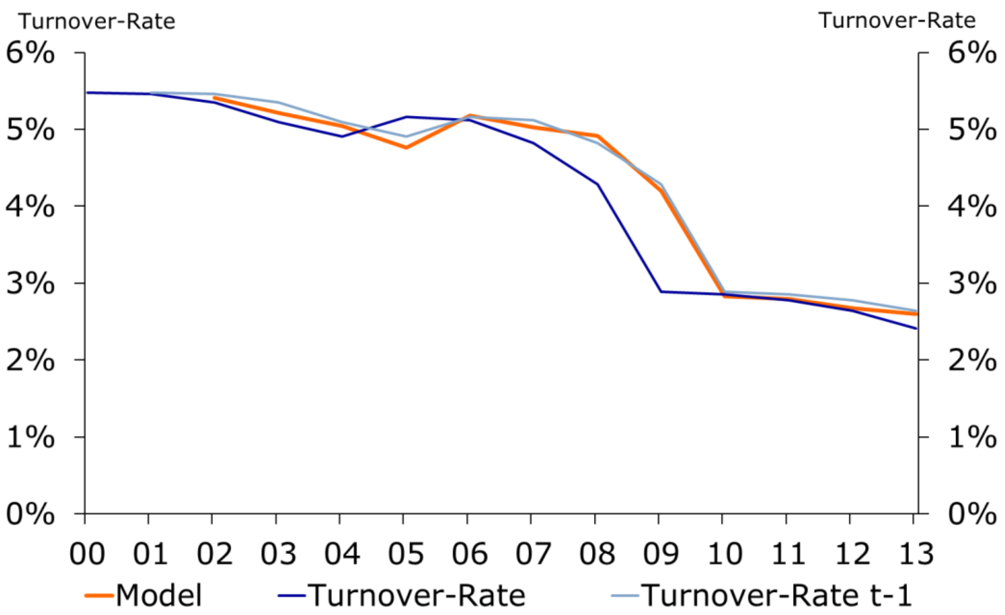
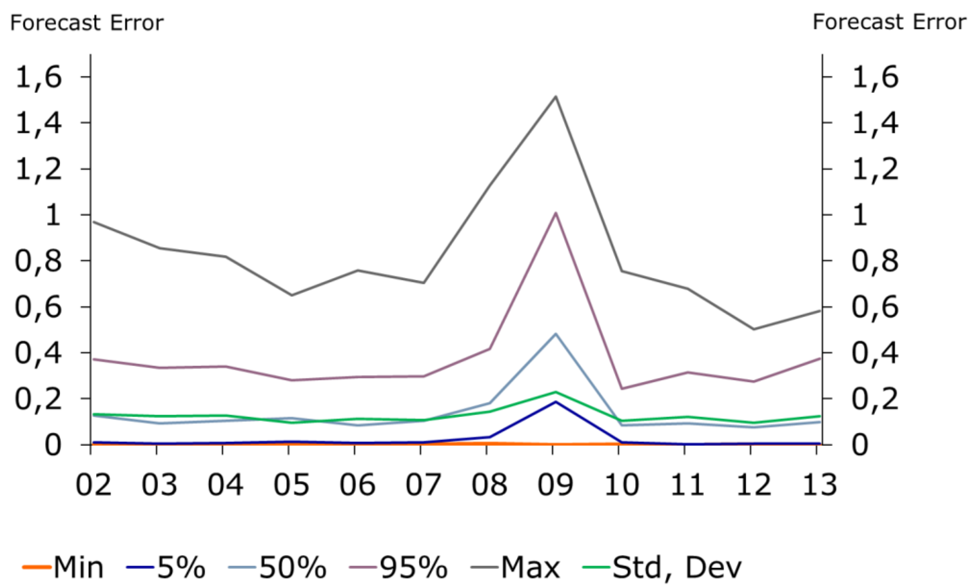


Figure 14: Forecast error of the model stated in levels throughout time



As Figure 14 shows, median forecast error is quite low (around 0.1 before and after 2009). Around 2009, the forecast error spikes to about 0.45. The median forecast error is in most years quite close to p5 and p95, which are around 0.007 and 0.3 respectively. This means that this model explains the turnover-rate quite well for most municipalities. There are also municipality clusters for which the turnover-rate is not predicted well, resulting in high maximum forecast errors of around 0.8.

All in all, results suggests that the chosen municipality economic and municipality demographic variables explain the turnover-rate in levels quite well. However, the deviation in turnover-rate from year to year is not very well explained. This can be seen for example around 2009, when the Netherlands endured a crisis. Around this time, the explanatory power of the model decreased significantly. Therefore, a model is needed to explain variations in the turnover-rate from year to year as well. Section 5.2 will elaborate on this.

5.2 Turnover-rate model in differences

As can be seen from Table 9, general model (2) and (3) are used to run 4 regressions. The used method is a panel ADL(1) with clustered standard errors, differenced time fixed effects and municipality/urbanization fixed effects. The lag of house prices and the lag of turnover rate were both instrumented using their second, third and fourth lag in levels. The first model includes (lags) of the turnover-rate, average house price, mortgage interest rates, and household income. To regression 3 and 4 employment is added. Overall, the model explains the turnover-rate in first differences quite well. The R-squared is 0.530 for the turnover-rate equation, and 0.542 for the house price equation. This number is inflated due to the time differenced fixed effects and the municipality/COROP cluster fixed effects. Without these fixed effects, the centered R-squared of the regression drops to 0.199 and 0.099 for the turnover-rate and house price equation respectively.

Coefficients in model (1) show some expected and unexpected results. Firstly, the lag of house prices price is significantly negative (-0.007) using a 90% confidence interval. This means that if house prices were to increase by 10% this year, then the overall turnover-rate is expected to decrease by 0.0007 in the next. This effect is very small. However, the results do match our stated **hypothesis 1** (the number of sold homes and the turnover-rate are significantly related). Even though we find a statistically significant coefficient for house prices, this negative coefficient is on the contrary to what was found when looking at correlations: the correlation coefficient between the lag of house prices and the turnover-rate was positive (and the correlation coefficient between house prices and the turnover-rate was significantly positive). An explanation to this discrepancy could be that, as explained by Dröes and Francke (2016), the turnover-rate and house prices are correlated positively because they have common factors that explain both house prices and the turnover-rate. As we took account of those variables in the regression, the positive correlation effect could disappear. The results are in line with former research like Dröes and Francke (2016) and Follain and Velz (1995) who also find a negative significant coefficient for the lag of house prices in the turnover-rate regression. Results are on the contrary to Clayton et al. (2008), who find significant positive coefficients for the first lag of house prices. As this research does not directly investigate the effect of a decrease and increase in price separately, there is no evidence to be found for either credit constraints (Stein, 1995; Follain & Velz 1995) or nominal loss aversion (Genesove & Mayer, 2001).

Table 9: Turnover-rate model in differences using entire dataset

	(1) Δ(Turnover-Rate)	(2) ΔLn(Transaction Price)	(3) Δ(Turnover-Rate)	(4) ΔLn(Transaction Price)
Δ(Turnover-Rate)$_{t-1}$	0.268***	0.785	0.269***	0.844
	(0.047)	(0.577)	(0.047)	(0.573)
ΔLn(Transaction Price)$_{t-1}$	-0.007*	0.167***	-0.007	0.170***
	(0.004)	(0.042)	(0.004)	(0.043)
Δ(Interest Rate)$_{t-1}$	-0.119**	-0.22	-0.122**	-0.378
	(0.049)	(0.431)	(0.056)	(0.494)
Δ(Interest Rate)$_{t-2}$	-1.070***	-2.770**	-1.074***	-2.916**
	(0.145)	(1.259)	(0.147)	(1.288)
ΔLn(Income)$_{t-1}$	0.020*	0.270**	0.020*	0.242*
	(0.010)	(0.137)	(0.010)	(0.135)
ΔLn(Employment)			0.0009	0.126
			(0.012)	(0.107)
N	1060	1060	1060	1060
Centered R^2	0.530	0.542	0.530	0.541
F	161.04***	338.54***	150.91***	307.42***
RMSE	0.004	0.038	0.004	0.038
Time Fixed effects	Yes	Yes	Yes	Yes
Cluster fixed effects	Yes	Yes	Yes	Yes

*Notes: Regressions used are panel ADL(1) with clustered standard errors. Both house prices and turnover-rates were instrumented using their second, third and fourth lag stated in levels using two stage GMM. Regressions are on the entire sample (2000-2013) and exclude outliers. Δ indicates that the first difference is taken. Ln indicates that the natural logarithm is taken. * Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level.*

The lag of the turnover-rate itself is found to be significantly positive in the turnover-rate regression (regression 1 and 3) using a 99% confidence interval. If the turnover-rate increases by 0.01 this year, this model predicts an increase of 0.003 the next. House prices have a significant positive coefficient in the house price regression (regression 2 and 4). If house prices increase by 1% this year, the model predicts house prices to increase by 0.2%

in the next. As is often observed, when the housing market is on the rise, it tends to do so for several periods. This could mean that perhaps both the turnover-rate and house prices are subject to momentum **(hypothesis 2)**. These results are in line with research like Dröes and Francke (2016), but only partially with Clayton et al. (2008) as they only find this momentum effect in the house price regression.

The lag of average household income is significantly positive using a 90% confidence level in regression 1 and 3. On average, this model predicts an increase of 0.002 in the turnover-rate when income has increased by 10%. It is perhaps not that surprising that the coefficient of income is small. As income is notorious for always increasing slightly but never decreasing, having a large positive coefficient would result in ever increasing turnover-rates. Even though the coefficient is small, this is evidence that perhaps large increases in income results in increasing turnover-rates **(hypothesis 3)**. An explanation to this could be that a higher income makes obtaining a mortgage easier. These findings are in line with findings of Dröes and Francke (2016) and Clayton et al. (2008).

Both the first and second lag of mortgage interest rate are found to have a significant negative effect on the turnover-rate. This means that the mortgage interest rate perhaps has a long lasting effect on the turnover-rate. For example, if the interest rate increases by 1 percentage this year, then we expect the turnover-rate to decrease by 0.001 next year, and with 0.011 the year after. These findings are in line with **hypothesis 4** (an increase in mortgage interest rates has a negative effect on the number of sold homes). An explanation for this effect could be that increasing mortgage interest rate mean a more difficult to obtain mortgage. This would in effect result in a decrease of transactions and turnover-rate. These findings are in line with findings of Dröes and Francke (2016) and Clayton et al. (2008).

Going from model 1 (excluding employment) to model 3 (including employment), coefficients do not change significantly. Actually, the only coefficient to change is the turnover-rate coefficient, which increases by 0.001. In addition, adding the logarithm of employment does not yield a significant coefficient. It is thus not surprising that the centered R-squared of this regression does not increase, indicating that employment does not add much explanatory power to the model. This could mean that changes in employment do not significantly affect the turnover-rate between 2000-2013 in the Netherlands. Therefore this research provides no such evidence that the turnover would increase if employment increases **(hypothesis 5)**.

All in all, the turnover-rate model in first differences explains variations in the turnover-rate fairly well. Combining these results with the turnover-rate model stated in levels, it could mean that municipality economic and municipality demographic variables explain the turnover-rate well in levels. Meanwhile, (multiple lags of) the turnover-rate, house prices, mortgage interest rates and household income are key determinants for *variations* of the turnover-rate in the Netherlands as a whole. This could mean that, to get a comprehensive view of what makes up the turnover-rate, we have to gain significant understanding of both models.

5.2.1 Turnover-rate model in differences- Urbanization model

Table 10 shows the model for 3 urbanization types. All regressions were set up in the same way as regressions in Table 9, adding household income and the second lag of household income to add more definition to the model. The model does not include COROP employment, as this yielded in non-significant coefficients in regressions 3 and 4 in Table 9. Due to a low number of observations, it is decided to combine both non-urban & low urban areas and high-urban & very-high urban areas together. All in all, the models based on urbanization levels explain variations in the differenced turnover-rate quite well. This results in centered R-squared values of 0.519, 0.535 and 0.480 for Non-/Low-urban, Moderate-urban and High-/Very-high urban areas respectively. Just as the turnover-rate model for the Netherlands as a whole, the centered R-squared measure is slightly inflated by the used fixed effects. However, not nearly as much as the model for the Netherlands as a whole. When the fixed effects are deleted, then the centered R-squared decreased to 0.387, 0.464 and increased to 0.507 for Non-/Low-urban, Moderate-urban and High-/Very-high urban areas respectively. Signs of coefficients are roughly in line with the whole sample model. Magnitude of these coefficients however differ greatly throughout urbanization levels.

First of all, it is hard to say whether house price changes affect the number of sold homes more in densely populated areas than in thinly populated areas **(hypothesis 6)**. Looking at Table 10, none of the house price coefficients are statistically significantly different from 0 in the turnover-rate equation. Therefore this research finds no evidence for a larger price effect in more urban areas (hypothesis 6)[4]. Secondly, the second lag of mortgage interest rates are significantly negative in all regressions. Meanwhile, the first lag is not significant in all urbanization regressions. This is on the contrary to the entire sample model in Table 9, where both lags were significantly negative. This effect is peculiar, as the dataset used for the whole sample returned significant estimates. The entire sample model is just an addition of non-, low, moderate-, high, and very-high urban areas.

Nevertheless, this does not mean that the regressions in Table 10 could not provide any interesting insights. In regression 9, the lag of turnover-rate is twice as big compared to regression 5 and 7. The same hold for the house price regression: the coefficient of the lag of house prices is more than twice as big in regression 9 compared to 6 (8 returns an

[4] One has to note here that big cities that the big cities like Utrecht and Amsterdam that were experiencing high growth in both turnover-rate and prices, were deleted through the outlier analysis. Therefore, those large cities were not able to drive the results. If they were not deleted from the sample, then perhaps the results could have been different.

insignificant coefficient). This could mean that high- and very-high urban areas are more heavily subject to momentum compared to non-, low- and moderate-urban areas.

Also interesting in regression 5 is the first lag of household income. This is significantly positive (0.015), meanwhile the second lag is significantly negative (-0.029). The negative coefficient is almost twice as great as the positive coefficient. This could be evidence that the income effect reverses over time. One year after income increases, the turnover-rate increases. The year after, this effect is completely reversed. This can be possibly be explained by the fact that households are not expected to move house again after they had just moved due to an increase in household income. Sadly enough, this effect is not easily compared to other urbanization levels. Only (very) weak evidence is found for this pattern in regression 9, where household income is positive (but not significant) meanwhile the first lag of income is significantly negative. Regression 7 on the other hand only has negative coefficients for the lags of income, only the second one being significant.

Table 10: Turnover-rate model in differences using urbanization levels

	Two stage GMM		Two stage GMM		Two stage GMM	
	(5)	(6)	(7)	(8)	(9)	(10)
	Non- and low-urban		Moderate urban		High- and Very-high urban	
	∆(Turnover-Rate)	∆Ln(Transaction Price)	∆(Turnover-Rate)	∆Ln(Transaction Price)	∆(Turnover-Rate)	∆Ln(Transaction Price)
∆(Turnover-Rate)t-1	0.238***	1.018	0.288***	0.930	0.573***	0.417
	(0.053)	(0.807)	(0.064)	(0.798)	(0.144)	(0.379)
∆Ln(Transaction Price)t-1	-0.003	0.143***	-0.010	0.046	-0.009	0.275**
	(0.005)	(0.052)	(0.007)	(0.061)	(.0155)	(0.117)
∆(Interest Rate)t-1	-0.065	-1.726*	-0.006	0.899	-0.002	-2.257**
	(0.066)	(0.954)	(0.106)	(1.111)	(0.191)	(-1.123)
∆(Interest Rate)t-2	-0.595***	-2.937*	-0.839***	-5.947**	-1.348***	-1.152
	(0.159)	(1.542)	(0.252)	(2.205)	(0.309)	(1.410)
∆Ln(Income)	-0.009	0.174	-0.007	-0.225	0.017	0.244
	(0.011)	(0.166)	(0.020)	(0.203)	(0.034)	(0.244)
∆Ln(Income)t-1	0.015**	0.196	-0.014	0.418**	-0.081***	0.478***
	(0.007)	(0.143)	(0.014)	(0.166)	(0.023)	(0.178)
∆Ln(Income)t-2	-0.029**	0.230*	-0.035**	-0.051	0.011	- 0.162
	(0.007)	(0.118)	(0.014)	(0.106)	(0.024)	(0.149)
N	550	550	300	300	300	300
Centered *R*2	0.519	0.511	0.535	0.575	0.480	0.667
F	102.49***	165.05***	73.97***	248.48***	267.79***	276.38***
RMSE	0.003	0.042	0.004	0.038	0.005	0.0295
Time Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cluster fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: Regressions used are panel ADL(1) with clustered standard errors. Both house prices and turnover-rates were instrumented using their second, third and fourth lag stated in levels using two stage GMM. Regressions are on the different urbanization levels indicated. The timespan is 2000-2013 and exclude outliers. ∆ indicates that the first difference is taken. Ln indicates that the natural logarithm is taken. * Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level.*

6. Conclusion

This study scrutinized the relationship between house prices and the number of sold homes for the Netherlands between 2000-2013. The number of sold homes is measured by the turnover-rate, which is defined as the number of sold homes divided by the owner occupied housing stock. By doing so, this research looked for an answer to the research question: ***which determinants explain the number of houses sold in the Netherlands between 2000-2013, and are there differences between densely and thinly populated areas?*** To my knowledge, such research has not yet been done on a municipality level.

Instead of using the number of sold homes in the Netherlands, this study estimated the turnover-rate. This is calculated as the number of sold homes divided by the owner occupied housing stock. As some municipalities naturally had more sold homes due to a larger housing stock, using this measure made municipalities more comparable.

To give more depth to this research, two models were utilized: one to estimate the turnover-rate in levels, and one in first differences. The model estimated in levels was done using simple OLS with clustered standard errors. Independent variables in this model were both municipality economic and municipality demographic. The model estimated in first differences is a panel ADL(1) model. Both house prices and the turnover-rate were instrumented using their second, third and fourth lag in levels. This model used both municipality economic and national data.

This research suggests that childbirth, household age (15-44), the number of people living within one household and household income are key determinants in explaining the turnover-rate stated in levels. Firstly, results suggest that the turnover-rate is generally higher in municipalities where childbirth is high and were mostly young households live (age 15-44). The latter is in line with expectations of Dröes and Francke (2016). Secondly, it is found that the turnover-rate of a municipality is generally lower when on average more people live within one household. Thirdly, it is found that household income became increasingly significant throughout time, having a positive impact on the turnover-rate. Coefficients are overall quite constant over time, but do experience some trends. This model is found to be not particularly suited for predicting the turnover-rate, as it does still not outperform the predictive powers of the lag turnover-rate itself.

This research suggests that (lags of) the turnover-rate, house prices, mortgage interest rates and household income might be key variables in explaining the turnover-rate stated in first differences. This research finds effect of house prices to be significantly negative. These results are in line with findings of Dröes and Francke (2016) and Follain and Velz (1995) but is on the contrary to what Clayton et al. (2008) found. When running regressions on urbanization level, the mentioned effect disappeared. In addition, no evidence could be found for either the credit constraint or loss aversion theory. This is because the effect of an increase/decrease in house prices is not separately investigated. Secondly, evidence is found for momentum in both the turnover-rate and house price regression, using a model for the Netherlands as a whole. This means that when house prices and turnover-rates start increasing, they tend to keep increasing for several periods. For the turnover-rate, this effect is found to be most pronounced in high- and very-high urban areas. For house prices no such effect has been found using regressions on urbanization level. These findings are in line with Dröes and Francke (2016), but only partially with Clayton et al. (2008) as they only find this momentum effect in the house price regression. Thirdly, interest rates seem to have a long lasting negative effect on the turnover-rate using a model for Netherlands as a whole. Dröes and Francke (2016) and Clayton et al. (2008) have similar results. In addition, it is found that the second lag of mortgage interest rates might influence High-/very-high urban areas more compared to Non- and low- urban areas. Lastly, this research finds household income to have a small positive effect on the turnover-rate of the Netherlands as a whole. These findings are in line with research like Dröes and Francke (2016) and Clayton et al. (2008). However, this effect disappears in most of the urbanization regressions.

The biggest limitation of the panel ADL(1) model is that it does not take account the interdependency of the turnover-rate and house prices. Research like Dröes and Francke (2016) find that, in their dataset, not taking account of this relationship results in significant biased estimates. In order to solve this, a bivariate PVAR(1) model could be used. This however resulted in an unstable model (results of said model can be found in **appendix G)**.

Another (solvable) downside to the panel ADL(1) model is that it does not take account of variables stated in levels. This research tries to solve this problem by estimating the turnover-rate in levels separately. Another way to solve said problems is to estimate a panel vector error correction model (VECM). This model allows to include variables stated in levels, and can take account of the turnover-rate house price relationship. In addition, it could provide extra insights, as it can make a distinction between long term and short term trends. In order to make a distinction between short- and long term trends, one needs a considerable amount of years to do significant statistical inference. Right now, this study utilizes 14 years of data. I believe this is not yet enough for the panel VECM model to work well. In the future this will inevitably become possible. Therefore, future research could utilize this model to explain more dynamics of the Dutch housing market.

Future research should also keep on investigating differences between municipalities. This research now investigated differences using urbanization. However, a study based on COROP areas is also imaginable. In addition, as the turnover-rate model already shows that coefficients trend over time, further research should investigate whether coefficients have changed significantly after the 2008 crisis.

Implications of this research are twofold. The turnover-rate in levels is well explained by municipality economic and municipality demographic variables. This implies that if we want to know what *really* decides how often people move, we need to go back to the people that make this housing market. Secondly, if we want to know what influences increases or decreases in the turnover-rate, we need to look at other variables. This research suggest that the increase or decrease in turnover-rate is explained by (multiple lags of) the turnover-rate, house prices, mortgage interest rates and household income. This however does not mean that all areas will respond in the same way. As regressions on urbanization level show: different areas in the Netherlands respond differently to different kind of shocks. Thus, if financial institutions want to get a comprehensive view about mortgage production and the associated risks, they would have to take into account those differences between municipalities too.

Bibliography

Aalders, R.,& van Dalen, P.P. (2016). Stijging Nederlandse Woningverkopen en huizenprijzen houdt aan. Retrieved from https://economie.rabobank.com/publicaties/2016/april/stijging-nederlandse-woningverkopen-en-huizenprijzen-houdt-aan/ on June 29, 2016.

Andrew, M., & Meen, G. (2003a). House price appreciation, transactions and structural change in the British housing market: a macroeconomic perspective. *Real Estate Economics*, *31*(1), 99-116.

Andrew, M., & Meen, G. (2003b). Housing transactions and the changing decisions of young households in Britain: the microeconomic evidence. *Real Estate Economics*, *31*(1), 117-138.

Banks, J., Blundell, R., & Oldfield, Z. (2004). House price volatility and housing ownership over the lifecycle.

Berkovec, J. A., & Goodman, J. L. (1996). Turnover as a measure of demand for existing homes. *Real Estate Economics*, *24*(4), 421-440.

Canova, F., & Ciccarelli, M. (2013). Panel Vector Autoregressive Models: A Survey☆☆ The views expressed in this article are those of the authors and do not necessarily reflect those of the ECB or the Eurosystem. *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims (Advances in Econometrics, Volume 32) Emerald Group Publishing Limited*, *32*, 205-246.

Carrillo, P. E., De Wit, E. R., & Larson, W. D. (2014). Can tightness in the housing market help predict subsequent home price appreciation? evidence from the us and the Netherlands. *Evidence from the US and the Netherlands (April 5, 2014)*.

CBS, 2015a. *Stedelijkheid van een Gebied.* Retrieved from http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?ConceptID=658 on March 9, 2015.

CBS, 2015b. *Gemeentelijke Indeling op 1 Januari 2015.* Retrieved from http://www.cbs.nl/nl-nl/menu/methoden/classificaties/overzicht/gemeentelijke-indeling/2015/default.html on March 14, 2015.

CBS, 2015c. Gemeente Indeling op Januari 2015. Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke indelingen/indeling%20per%20jaar/gemeentelijke-indeling-op-1-januari-2015 on April 22, 2015.

CBS, 2015d. Gemeente Indeling op Januari 2014. Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen/indeling%20per%20jaar/gemeentelijke-indeling-op-1-januari-2016/gemeentelijke-indeling-op-1-januari-2014 on April 22, 2015.

CBS, 2015e. Gemeente Indeling op Januari 2013. Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/overig/gemeentelijke-indelingen/indeling%20per%20jaar/gemeentelijke-indeling-op-1-januari-2016/gemeentelijke-indeling-op-1-januari-2013 on April 22, 2015.

CBS, 2015f. Landelijk dekkende indelingen. Retrieved from https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/gemeente/gemeenten-en-regionale-indelingen/landelijk-dekkende-indelingen on May 6, 2016.

CBS, 2015g. Regionaal Inkomensonderzoek (RIO). Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/regionaal-inkomensonderzoek--rio-- on June 24, 2016.

CBS, 2015h. Bevolking en bevolkingsontwikkeling; per maand, kwartaal en jaar. Retrieved from http://statline.cbs.nl/StatWeb/publication/?PA=37943ned on 24 june 2016.

CBS, 2015i. Gemiddeld inkomen; particuliere huishoudens naar diverse kenmerken. Retrieved from http://statline.cbs.nl/statweb/publication/?dm=slnl&pa=708 43ned&d1=a&d2=0&d3=0&d4=a&hdr=g1,g2,g3&stb=t&vw=t on 24 June, 2016.

Clayton, J., Miller, N., & Peng, L. (2010). Price-volume correlation in the housing market: causality and co-movements. *The Journal of Real Estate Finance and Economics*, *40*(1), 14-40.

de Vries, P. (2014). Herstelvermogen van huishoudens die onder water staan. Retrieved from https://economie.rabobank.com/publicaties/2014/juni/herstelvermogen-van-huishoudens-die-onder-water-staan/ on April 15, 2015.

de Vries, P., & Boelhouwer, P. (2009). Equilibrium between interest payments and income in the housing market. *Journal of Housing and the Built Environment*, *24*(1), 19-29.

de Wit, E. R., Englund, P., & Francke, M. K. (2013). Price and transaction volume in the Dutch housing market. *Regional Science and Urban Economics*,*43*(2), 220-241.

DiPasquale, D., & Wheaton, W. C. (1994). Housing market dynamics and the future of housing prices. *Journal of urban economics*, *35*(1), 1-27.

Dröes, M. I., & Francke, M. (2016). *The Price-Turnover Relationship in European Housing Markets* (No. eres2015_295). European Real Estate Society (ERES).

Dua, P., & Miller, S. M. (1996). Forecasting Connecticut home sales in a BVAR framework using coincident and leading indexes. *The Journal of Real Estate Finance and Economics*, *13*(3), 219-235.

Dua, P., & Smyth, D. J. (1995). Forecasting US home sales using BVAR models and survey data on households' buying attitudes for homes. *Journal of forecasting*, *14*(3), 217-227.

Fisher, J., Gatzlaff, D., Geltner, D., & Haurin, D. (2004). An analysis of the determinants of transaction commercial real estate investment property. *Real*

Estate Economics, *32*(2), 239-264.

Follain, J. R., & Velz, O. T. (1995). Incorporating the number of existing home sales into a structural model of the market for owner-occupied housing. *Journal of Housing Economics*, *4*(2), 93-117.

Francke, M. (2010). *Casametrie: de kunst van het modelleren en het voorspellen van de marktwaarde van woningen* (Vol. 353). Amsterdam University Press.

Francke, M.K., S. Vujić, & G.A. Vos, 'Forecasting House Prices.', Mimeo, Universiteit van Amsterdam, 2010 (this is an update of: 'Evaluation of House Price Models Using an ECM Approach.' ERES Conference, Stokholm, 2009).

Genesove, D., & Han, L. (2012). Search and matching in the housing market. *Journal of Urban Economics*, *72*(1), 31-45.

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

Han, L. (2013). Understanding the puzzling risk-return relationship for housing. *Review of Financial Studies*, *26*(4), 877-928.

Hort, K. (2000). Prices and turnover in the market for owner-occupied homes. *Regional Science and Urban Economics*, *30*(1), 99-119.

Meen, G. (1999). Regional house prices and the ripple effect: a new interpretation. *Housing studies*, *14*(6), 733-753.

Oevering, F. (2014). *Visie op de Woningmarkt.* Retrieved from https://economie.rabobank.com/Documents/2014/SP20140129_Visie_op_de_woningmarkt_Frits_Oevering.pdf on February 15, 2015.

Sinai, T., Souleles, N.S., 2005. Owner-Occupied Housing as a Hedge against Rent Risk, 7 Quarterly Journal of Economics 120, 763-789.

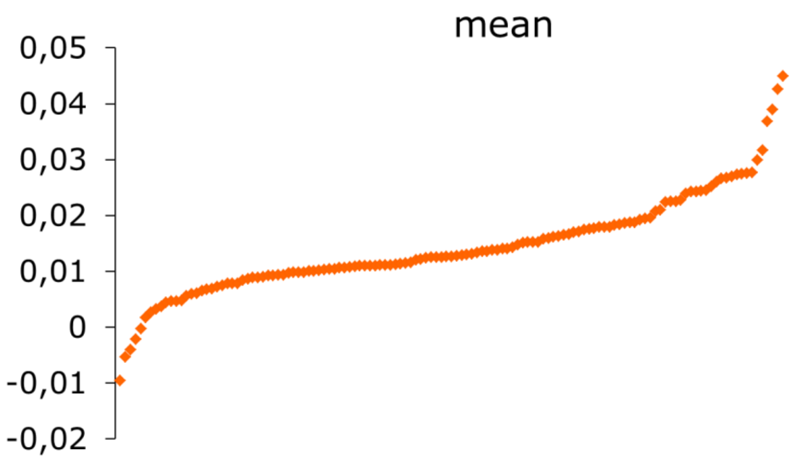
Stein, J.C., 1995. Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects, Quarterly Journal of Economics 110, 379-406.

Van Dijk, D. W., & Francke, M. (2015). Internet search behavior, liquidity and prices in the housing market.

Appendix A: Outlier analysis

This appendix shows figures of mean, median, standard deviation, minimum, maximum and autocorrelation statistics of both the turnover-rate stated and the natural logarithm of average house prices stated in first differences. This is done in order to find municipalities that behave differently from others. This analysis is done using clustered (COROP/urbanization) municipalities. How those municipalities were clustered can be found in **appendix C.**

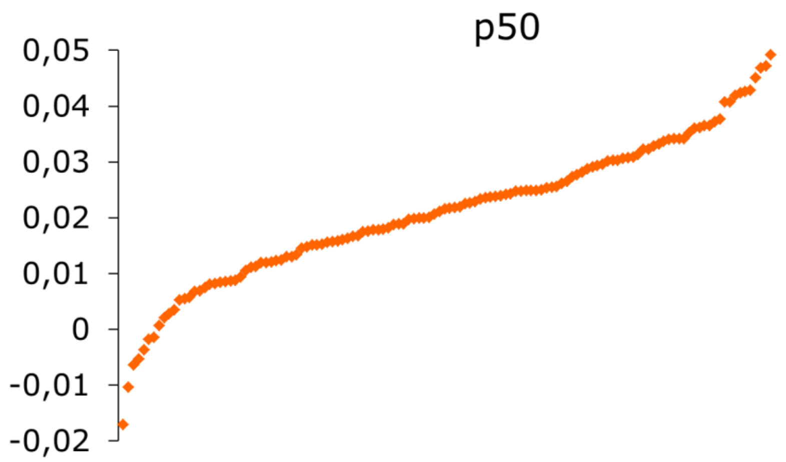
Figure 1: Mean ∆Ln(Transaction Price)



Outliers High: ID=21; 55; 116; 122

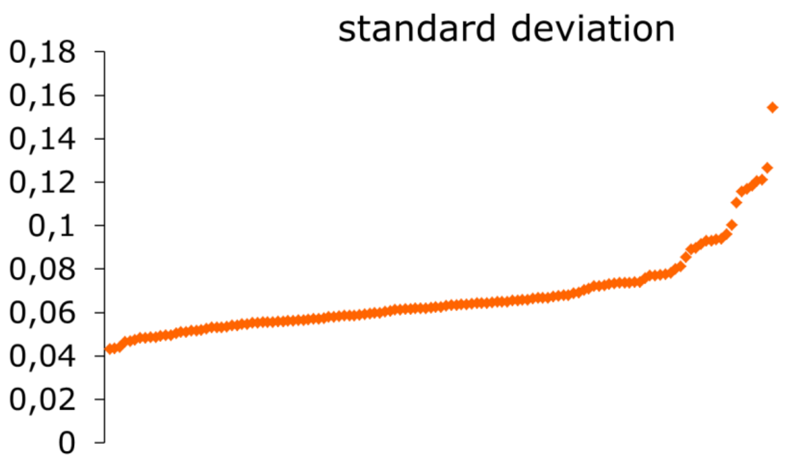
Outliers Low: ID= 26

Figure 2: Median ∆Ln(Transaction Price)



Outliers Low: ID= 82

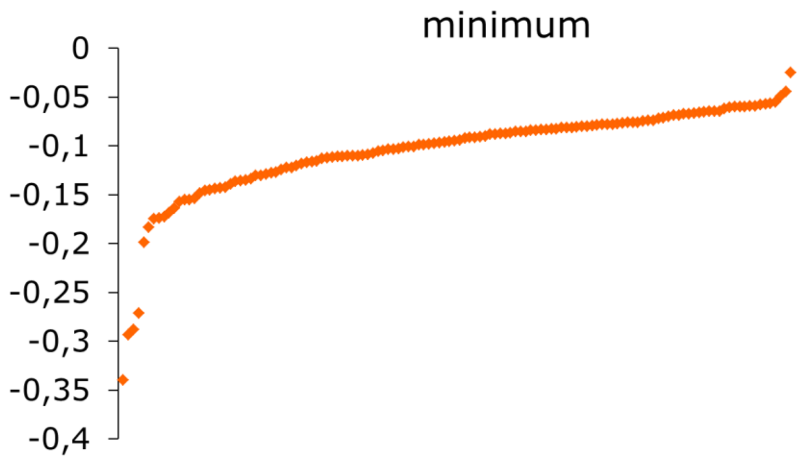
Figure 3: Standard Deviation ∆Ln(Transaction Price)



Outliers High: ID= 10; 17; 18; 19; 23; 26; 27; 50; 80[5]

[5] Please note that ID 10, 17, 26, 27, 55 were also formerly deleted due to outliers is one of the other summary statistics of ∆Ln(Transaction Price) and/or ∆(Turnover)

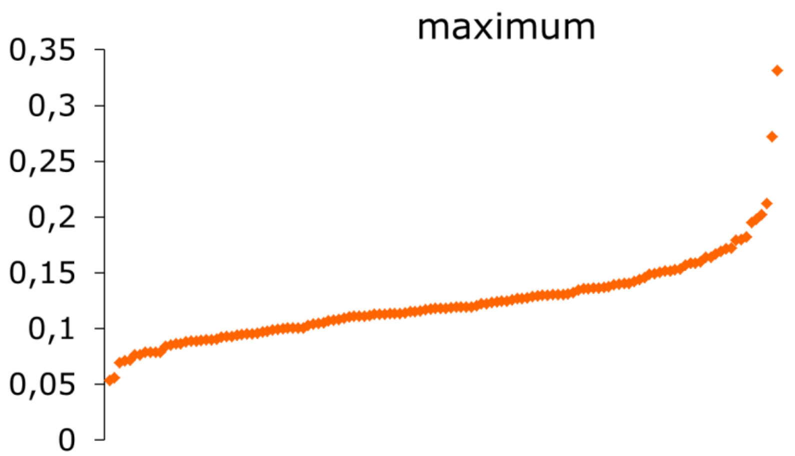
Figure 4: Minimum ΔLn(Transaction Price)



Outliers Low: ID= 55[5]

Outliers high: ID= 10; 17; 26; 27[5]

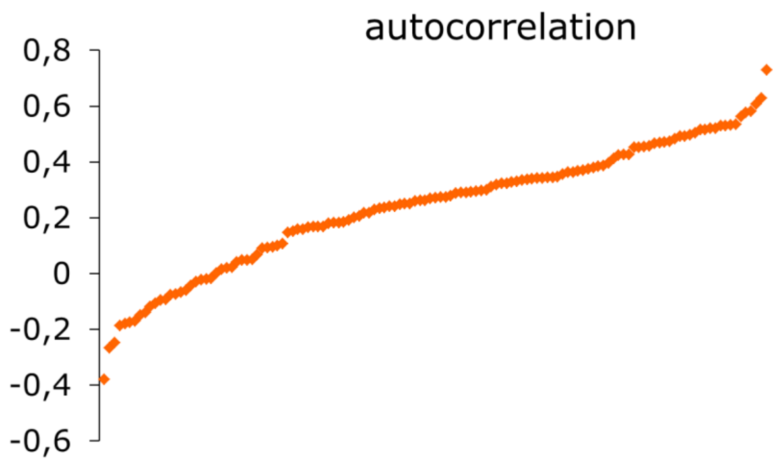
Figure 5: Maximum ΔLn(Transaction Price)



Outliers High: ID= 17; 23[6]

[6] Please note that ID 10, 13 , 17, 23, 50 and 116 were also formerly deleted due to outliers is one of the other summary statistics of ΔLn(Transaction Price) and/or Δ(Turnover)

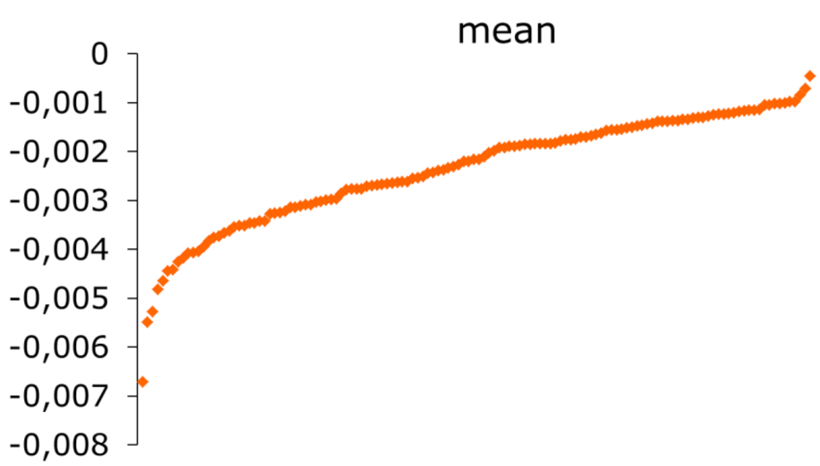
Figure 6: Autocorrelation ∆Ln(Transaction Price)



Outliers Low: ID= 10; 13; 50[6]

Outliers High: ID= 116[6]

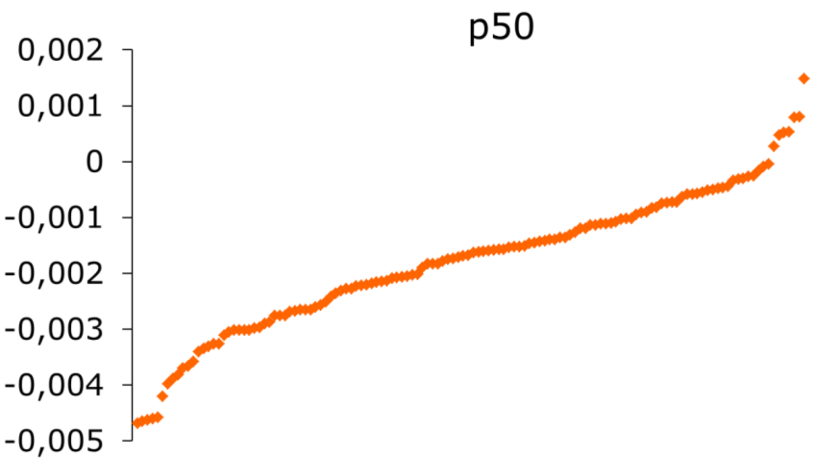
Figure 7: Mean ∆Turnover-Rate



Outliers Low: ID=98 ;123 ;130[7]

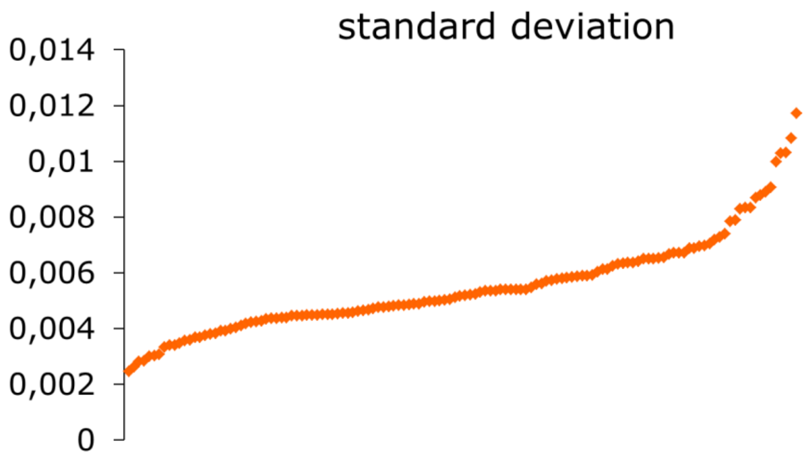
[7] Please note that ID 8, 26, 27, 80 and 98 was also formerly deleted due to outliers is one of the other summary statistics of ΔLn(Transaction Price) and/or Δ(Turnover)

Figure 8: Median ∆Turnover-Rate



Outliers High: ID=80[7]

Figure 9: Standard Deviation ∆Turnover-Rate

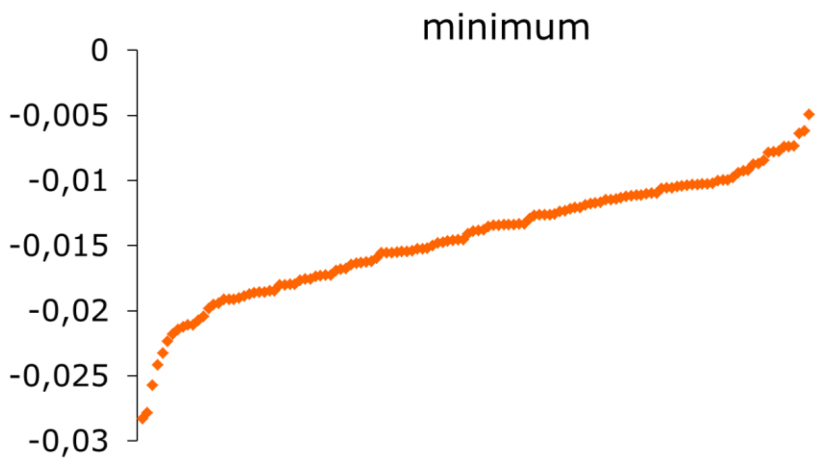


Outliers High: ID= 18; 19; 97; 109; 126[8]

---

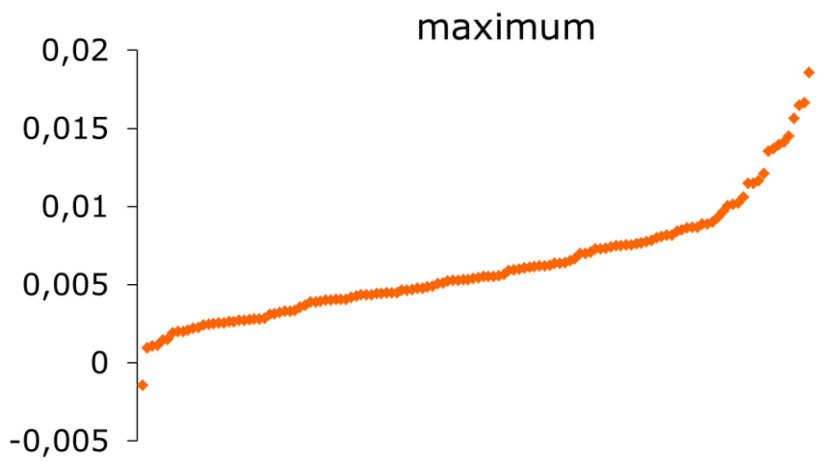
[8] Please note that ID 18, 19,48, 97 and 126 were also formerly deleted due to outliers is one of the other summary statistics of ΔLn(Transaction Price) and/or Δ(Turnover)

Figure 10: Minimum ∆Turnover-Rate



Outliers low: ID= 48; 109; 126[8]

Figure 11: Maximum ∆Turnover-Rate

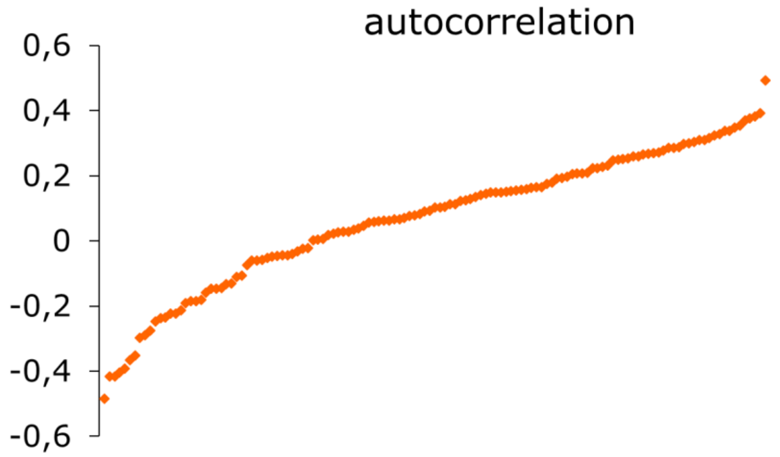


Outliers High: ID= 10;18; 19; 56; 97[9]

Outliers low: ID= 130[9]

[9] Please note that ID 10,13, 18, 19, 56, 80, 97, 126, 130 were also formerly deleted due to outliers is one of the other summary statistics of ∆Ln(Transaction Price) and/or ∆(Turnover)

Figure 12: Autocorrelation ∆Turnover



Outliers High: ID= 112; 126[9]

Outliers low: ID= 13; 33[9]

Conclusion:

Given this outlier analysis, ID clusters 10, 13, 17, 18, 19, 21,23, 26 , 27, 33, 48, 50, 55; 56, 82, 80, 97, 98, 109,112, 116, 122, 123,126,130 were deleted from the sample.

Appendix B: Conversion 2012-2015 municipality data

According to CBS 2012 division, The Netherlands holds 415 municipalities. In 2015 this has decreased to 393 municipalities. All in all, 5 municipalities have been added, 27 have disappeared and 4 municipalities have been renamed (CBS, 2015c; CBS,2015d; CBS,2015e).

Municipalities Renamed:

To fit the 2015 division of municipalities, several municipalities had to be renamed:
- 's-Gravenhage renamed to Den Haag
- Haarlemmerliede en Spaarnwoude renamed to Haarlemmerliede c.a.
- Kollumerland en Nieuwkruisland renamed to Kollumerland c.a.
- Nuenen, Gerwen en Nederwetten renamed to Nuenen c.a.

Municipalities created:

From 2012-2015, some municipalities were created by adding different municipalities together. So, the housing stock of the 2012 division municipalities had to be added together to create the 2015 definition municipality.

- De Friese Meren created. Created by adding Gaasterlân-Sleat, Lemsterland, Skarsterlân and Boarnsterhim together.
- Goeree-Overflakkee created. Created by adding Dirksland Goedereede Middelharnis Oostflakkee together.
- Krimpenerwaard created. Created by adding Bergambacht, Nederlek, Ouderkerk, Schoonhoven and Vlist together.
- Molenwaard created. Created by adding Graafstroom, Liesveld and Nieuw-Lekkerland together.
- Nissewaard created. Created by adding Bernisse and Spijkernisse.

Added to existing municipalities:

From 2012-2015, some municipalities were added to excising municipalities. In order to do this, the housing stock of those municipalities were added to the already existing municipalities.

- To Alphen aan den Rijn is added: Boskoop and Rijnwoude
- To Alkmaar are added: Graft-De-Rijp and Schermer
- To Schagen are added: Harenkarspel and Zijpe
- To Groesbeek is added: Millingen aan de Rijn and Ubbergen

Special:

One municipality was split and added to two different existing municipalities. In order to comply with the 2015 CBS division of municipalities, the housing stock of this municipality had to be split and added to those municipalities.

- Maasdonk is split into ‘s-Hertogenbosch and Oss. CBS does not provide any reasoning to how Maasdonk was split. Therefore, the assumption was made that housing stock has been split equally between both municipalities.

Missing data- Extrapolation:

In order to fit the 2000-2015 timeframe, housing stock on all municipalities have to be extrapolated. However, for some municipalities an even less large timeframe is available on housing stock. Therefore these municipalities have a housing stock that needs to be extrapolated on a larger timeframe. Those municipalities are listed below. This list shows for what years the municipality has missing data points.

- Bodegraven- Reeuwijk data missing 2006-2010.
- Dantumadiel data missing 2006-2008.
- Eijsden-Margraten data missing 2006-2010.
- Kaag en Braassem data missing 2006-2008.
- Koggenland data missing 2006.
- Lansingerland data missing 2006.
- Leudal data missing 2006.
- Maasgouw data missing 2006.
- Menameradiel data missing 2006-2010.
- Oldambt data missing 2006-2009.
- Oost Gelre data missing 2006.
- Peel en Maas data missing 2006-2009.
- Stichtse Vecht data missing 2006-2010.
- Súdwest Fryslân data missing 2006-2010.
- Zuidplas data missing 2006-2009.

Municipalities deleted:

- Hollands Kroon deleted due to having only one observation in 2012 (extrapolation not possible)
- Binnenmaas deleted due to having too high growth in a short period of time (extrapolation returns negative values)

- Medemblik deleted due to having too high growth in a short period of time (extrapolation returns negative values)
- Nieuwkoop deleted due to having too high growth in a short period of time (extrapolation returns negative values)
- Roerdalen deleted due to having too high growth in a short period of time (extrapolation returns negative values)
- Roermond deleted due to having too high growth in a short period of time (extrapolation returns negative values)

Appendix C: Merged municipalities

Table 1: Merged municipalities

ID	Municipalities in cluster
1	Bellingwedde Vlagtwedde Menterwolde
2	Loppersum
3	Eemsmond Ten Boer Grootegast Winsum Slochteren De Marne Zuidhorn Marum
4	Menameradiel Vlieland Kollumerlandc.a. Achtkarspelen Ameland Schiermonnikoog Ferwerderadiel Littenseradiel Terschelling Het Bildt Tytsjerksteradiel Dantumadiel Dongeradeel
5	Ooststellingwerf Opsterland
6	Tynaarlo Midden-Drenthe Aa en Hunze
7	Coevorden Borger-Odoorn
8	De Wolden Westerveld
9	Ommen Staphorst Dalfsen
10	Olst-Wijhe
11	Tubbergen Dinkelland
12	Bronckhorst
13	Winterswijk Doetinchem Zutphen
14	Neder-Betuwe Neerijnen Maasdriel Lingewaal West Maas en Waal Buren
15	Lopik Renswoude
16	Drechterland Opmeer Texel Koggenland
17	Zeevang
18	Muiden
19	Korendijk
20	Zederik Giessenlanden Molenwaard
21	Sluis Hulst
22	Noord-Beveland Reimerswaal Tholen Schouwen-Duiveland Borsele Veere
23	Zundert
24	Aalburg Alphen-Chaam Woudrichem Baarle-Nassau
25	Sint Anthonis Haaren Landerd Boekel Millen Sint Hubert
26	Bergeijk
27	Mook en Middelaar Bergen(L.)

28	Leudal Maasgouw
29	Gulpen-Wittem Onderbanken Eijsden-Margraten Schinnen Nuth
30	Stadskanaal Oldambt Pekela Veendam
31	Appingedam Delfzijl
32	Leek Haren Bedum
33	Leeuwarderadeel Franekeradeel Harlingen
34	De Friese Meren Súdwest Fryslân
35	Weststellingwerf
36	Noordenveld
37	Emmen
38	Zwartewaterland Steenwijkerland Hardenberg
39	Raalte
40	Hof van Twente Haaksbergen Wierden Hellendoorn Losser Twenterand
41	Ermelo Nunspeet Voorst Oldebroek Heerde Hattem Scherpenzeel Putten Nijkerk Elburg Barneveld Epe
42	Aalten Lochem Berkelland Oude IJsselstreek Montferland Brummen Oost Gelre
43	Rozendaal Heumen Overbetuwe Beuningen Lingewaard Renkum Doesburg Druten Groesbeek
44	Zaltbommel Geldermalsen
45	Montfoort Rhenen Eemnes Bunnik Woudenberg Vianen Oudewater De Ronde Venen Utrechtse Heuvelrug
46	Schagen
47	Bergen(NH.) Langedijk
48	Haarlemmerliedec.a. Bloemendaal
49	Beemster Oostzaan Aalsmeer Landsmeer Waterland
50	Wijdemeren Blaricum
51	Zoeterwoude Kaag en Braassem
52	Krimpenerwaard
53	Westvoorne Goeree-Overflakkee Albrandswaard Cromstrijen Strijen Brielle
54	Hardinxveld-Giessendam
55	Terneuzen
56	Kapelle
57	Steenbergen Woensdrecht Rucphen Drimmelen Moerdijk Halderberge

58	Oisterwijk Werkendam Loon op Zand Hilvarenbeek
59	Cuijk Heusden Bernheze Veghel Sint-Michielsgestel Sint-Oedenrode Schijndel Grave Boxmeer
60	Waalre Reusel-De Mierden Deurne Laarbeek Son en Breugel Gemert-Bakel Oirschot Someren Bladel Cranendonck Asten Heeze-Leende Eersel
61	Gennep Peel en Maas Horst aan de Maas Venray Beesel
62	Nederweert Echt-Susteren
63	Vaals Simpelveld Beek Voerendaal Stein Valkenburg aan de Geul Meerssen
64	Zeewolde Dronten Urk Noordoostpolder
65	Hoogezand-Sappemeer
66	Smallingerland Heerenveen
67	Assen
68	Meppel Hoogeveen
69	Kampen
70	Rijssen-Holten Borne Almelo Oldenzaal
71	Ermelo Nunspeet Voorst Oldebroek Heerde Hattem Scherpenzeel Putten Nijkerk Elburg Barneveld Epe
72	Winterswijk Doetinchem Zutphen
73	Duiven Zevenaar Wijchen Westervoort Rheden
74	Culemborg Tiel
75	Wijk bij Duurstede Zeist Baarn Soest De Bilt Stichtse Vecht Woerden Houten Leusden Bunschoten
76	Stede Broec Enkhuizen
77	Heiloo
78	Uitgeest Castricum
79	Zandvoort
80	Wormerland
81	Uithoorn Edam-Volendam Ouder-Amstel Haarlemmermeer
82	Laren Naarden
83	Teylingen Noordwijk Hillegom Lisse Noordwijkerhout
84	Pijnacker-Nootdorp Wassenaar
85	Midden-Delfland Westland
86	Waddinxveen Bodegraven-Reeuwijk
87	Zuidplas Oud-Beijerland Lansingerland

88	Leerdam Alblasserdam
89	Goes
90	Geertruidenberg Oosterhout Roosendaal
91	Waalwijk Gilze en Rijen Dongen Goirle
92	Uden Vught Boxtel Oss
93	Best Nuenenc.a. Valkenswaard Geldrop-Mierlo
94	Venlo
95	Weert
96	Sittard-Geleen Landgraaf Kerkrade
97	Lelystad
98	Leeuwarden
99	Zwolle
100	Deventer
101	Enschede Hengelo
102	Apeldoorn Wageningen
103	Nijmegen Arnhem
104	Veenendaal Nieuwegein IJsselstein Amersfoort
105	Hoorn Den Helder
106	Alkmaar Heerhugowaard
107	Velsen Heemskerk
108	Heemstede
109	Zaanstad
110	Diemen Amstelveen Purmerend
111	Hilversum Huizen Weesp Bussum
112	Voorschoten Oegstgeest Katwijk Leiderdorp
113	Alphen aan den Rijn Gouda
114	Krimpen aan den IJssel Maassluis Nissewaard Capelle aan den IJssel Barendrecht Hellevoetsluis Ridderkerk
115	Papendrecht Gorinchem Dordrecht Zwijndrecht Hendrik-Ido-Ambacht Sliedrecht
116	Vlissingen Middelburg
117	Breda Bergen op Zoom Etten-Leur
118	‘S-hertogenbosch

119	Eindhoven Veldhoven Helmond
120	Maastricht Brunssum Heerlen
121	Almere
122	Groningen
123	Utrecht
124	Beverwijk
125	Haarlem
126	Amsterdam
127	Leiden
128	Leidschendam-Voorburg Rijswijk Den Haag Zoetermeer
129	Delft
130	Rotterdam Schiedam Vlaardingen
131	Tilburg

Appendix D: Descriptive statistics for variables stated in first differences

Table 1: Descriptive Statistics of variables in first differences (Annual data 2000-2013)

	Mean	Median	N	Std. Dev	Min	Max
National Data						
∆ Mortgage Interest Rate (%)	-0.17	-0.33	13	0.45	-0.83	0.62
Municipality Economic						
∆(Turnover-Rate) (%)	-0.22	-0.16	1378	0.52	-2.41	1.41
∆Ln(Transaction Price) (€)	0.01	0.02	1378	0.06	-0.18	0.21
Ln(Employment)	12.02	12.00	1484	0.76	9.80	13.65
∆Ln(Employment)	0.07	0.005	1378	0.01	-0.04	0.07
Ln(Household Income)	10.38	10.38	1484	0.13	10.03	10.80
∆Ln (Household Income) (€)	0.02	0.01	1378	0.03	-0.04	0.13
Municipality Demographic						
Ln(Childbirth)	6.93	7.04	1484	0.84	4.53	9.18

Notes: This Table shows descriptive statistics for the transformed variables used in this research. They were divided into three categories: National-, municipality demographic-, and municipality economic data. If data was not transformed, then it is not reported in this Table. ∆ indicates that the first difference is taken. Ln indicates that the natural logarithm is taken. All data shown is shown without outliers from 2000-2013.

Appendix E: Descriptive statistics for different types of homes

Surprising is the difference between median and mean values of house sales per house type; the mean number of sold homes is significantly more than the median. This indicates that a municipalities' housing stock hold many the same types of houses. As it is the case that very-high urban areas hold more apartments than do less urban areas, it explains why house prices are the lowest in very-high urban areas, even though it is likely that house prices per m^2 are the most expensive in that area. All in all, this means that looking only at average house prices could give a distorted image throughout urbanization levels. Therefore, house price increases are likely to present a better image.

Table 1: Number of home sales and average house prices per house type

	Mean	Median	N	Std. Dev	Min	Max
Home Sales						
- Apartment	283	98	1484	753.50	0	9421
- End-of-terrace	178	143	1484	148.52	1	808
- Detached house	158	112	1484	138.09	0	785
- Semidetached house	154	114	1484	126.56	4	657
- Mid-terrace	440	308.5	1484	429.40	1	2168
House Price						
- Apartment *(Thousand €)*	168	167	1470	36	59	293
- End-of-terrace *(Thousand €)*	211	204	1484	60	77	541
- Detached house *(Thousand €)*	397	367	1483	154	125	1394
- Semidetached house *(Thousand €)*	263	243	1484	94	88	759
- Mid-terrace *(Thousand €)*	195	191	1484	51	76	482

Notes: This Table shows descriptive statistics for 5 types of houses throughout the Netherlands (2000-2013). The data is in a yearly format with the timespan 2000-2013. When there are no sold homes for one type of house in a certain year within a cluster, then the number of home sales is set to 0. The average transaction price is then set to missing.

Appendix F: Turnover-rate model in levels

Table 1: Demographic regressions

Date	(1) turnover 2001	(2) turnover 2002	(3) turnover 2003	(4) turnover 2004	(5) turnover 2005	(6) turnover 2006	(7) turnover 2007	(8) turnover 2008	(9) turnover 2009	(10) turnover 2010	(11) turnover 2011	(12) turnover 2012	(13) turnover 2013
Age1544t-1	0.100***	0.085***	0.080***	0.060***	0.095***	0.078***	0.073***	0.084***	0.054***	0.040***	0.040***	0.039***	0.031***
	(0.0261)	(0.023)	(0.021)	(0.020)	(0.017)	(0.016)	(0.020)	(0.014)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
Ln(Income)t-1	0.014	0.011	-0.004	0.025**	0.019	0.028***	0.027***	0.023***	0.018***	0.021***	0.025***	0.020***	0.022***
	(0.013)	(0.012)	(0.009)	(0.012)	(0.013)	(0.011)	(0.009)	(0.007)	(0.005)	(0.006)	(0.005)	(0.004)	(0.006)
Ln(ChildBirth)t-1	0.004***	0.005***	0.004**	0.004***	0.004***	0.004***	0.003*	0.002	0.002**	0.002***	0.0007	0.0009	0.001**
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0009)	(0.0006)	(0.0008)	(0.0006)	(0.0005)
People per householdt-1	-0.042***	-0.041***	-0.038***	-0.041***	-0.039***	-0.039***	-0.035***	-0.027***	-0.021***	-0.020***	-0.017***	-0.013***	-0.012***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Ln(Employment)t-1	-0.002	-0.002	-0.0004	-0.002	-0.002	-0.001	-0.002	-0.0008	-0.001	-0.001*	-0.001	-0.0007	-0.0008
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.0008)	(0.0006)	(0.0007)	(0.0006)	(0.0005)
Constant	-0.030	-0.001	0.130	-0.135	-0.093	-0.188*	-0.174**	-0.168***	-0.131***	-0.156***	-0.202***	-0.167***	-0.191***
	(0.127)	(0.112)	(0.085)	(0.116)	(0.117)	(0.097)	(0.080)	(0.063)	(0.049)	(0.053)	(0.046)	(0.042)	(0.052)
N	106	106	106	106	106	106	106	106	106	106	106	106	106
adj. *R*2	0.670	0.664	0.728	0.652	0.693	0.702	0.670	0.674	0.684	0.726	0.600	0.606	0.553
F	34.08***	31.12***	40.51***	27.72***	28.28***	33.32***	22.43***	47.33***	47.66***	44.50***	29.98***	17.15***	16.04***
RMSE	0.009	0.009	0.007	0.008	0.008	0.007	0.007	0.006	0.004	0.003	0.004	0.003	0.003

*Notes: Regressions used are simple OLS with clustered standard errors. Regressions are on the entire sample(2000-2013) (excluding outliers). Ln indicates that the natural logarithm is taken * Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level.*

Appendix G: Turnover-rate model in differences- Bivariate PVAR(1)

Another way to check our stated hypotheses is by running a bivariate PVAR(1) model. To ensure a good comparison between models, they are constructed as similarly as possible. The difference between the panel ADL model and the bivariate PVAR model is that the latter does not include cluster fixed effects. This is because this method does not directly allow putting 130 dummies (the number of clusters used minus 1) into the regression to take account of cluster fixed effects. Instead, all variables are demeaned. This means that for each variable, the mean of its municipality cluster is deducted. To ensure a little bit more stability within the model, house prices and the turnover-rate are instrumented on their second, third, fourth and fifth lag stated in levels. This means that one lag is added.

All in all, the panel ADL(1) model in presented in Table 11 looks somewhat similar to the bivariate PVAR(1). Signs and significance are in most cases similar to the panel ADL(1) model. The magnitude of those coefficients however differ significantly from one model to the next.

There are similarities between both models. The lag of transaction prices is again (in line with expectations) significant **(hypothesis 1).** The magnitude of the coefficient does vary considerably. In the panel ADL(1) model the coefficient of house prices was -0.007. This was -0.023 in the bivariate PVAR(1) model. The first and second lag of interest rates are negative and statistically significant **(hypothesis 4).** The coefficient of the first lag of mortgage interest rates is very similar. In the panel ADL(1) model this was -0.119, and in the bivariate PVAR(1) model this was -0.200. The coefficient of the second lag of mortgage interest rates differ substantially. In the panel ADL(1) model this was -1.070, meanwhile in the bivariate PVAR(1) model this was -0.307.

There are some dissimilarities as well. First of all, The lag of the turnover-rate in regression 1 is positive but insignificant. This is against expectations stated by **hypothesis 5.** The magnitude of the coefficient is very different as well. In the panel ADL(1) model this was 0.268, meanwhile in the bivariate PVAR(1) model this is only 0.031. The lag of house prices in the house price regression (regression 2) is also found to be insignificantly negative. This is again on the contrary to the panel ADL(1) model where this was positively significant. In the panel ADL(1) model this was 0.167, meanwhile in the bivariate PVAR(1) model this is only -0.039. This is again against expectations stated by **hypothesis 5.** The lag of household income is positive, but only significant in regression 3 **(hypothesis 3).** Employment is significant in regression 3 but is against expectations significantly negative **(hypothesis 2)**

For the entire sample bivariate PVAR(1) model, Granger causality tests were done. Results can be found in Table 2. Results suggest that house prices granger cause the turnover-rate, but the turnover-rate does not Granger cause house prices.

Table 1: Whole model (using a bivariate PVAR(1) regression)

	(1)		(2)	
	∆(Turnover-Rate)	∆Ln(Transaction Price)	∆(Turnover-Rate)	∆Ln(Transaction Price)
∆(Turnover-Rate)t-1	0.031	0.619	0.022	0.647
	(0.083)	(1.199)	(0.083)	(1.191)
∆Ln(Transaction Price)t-1	-0.023**	-0.039	-0.022**	-0.037
	(0.010)	(0.127)	(0.010)	(0.127)
∆(Interest Rate)t-1	-0.200***	-0.709	-0.173***	-0.856
	(0.064)	(0.006)	(0.067)	(0.007)
∆(Interest Rate)t-2	-0.307***	-3.33***	-0.312***	-3.29***
	(0.064)	(0.667)	(0.064)	(0.665)
∆Ln(Income)t-1	0.018	0.247*	0.020*	0.237*
	(0.011)	(0.144)	(0.011)	(0.143)
∆Ln(Employment)			-0.021**	0.107
			(0.011)	(0.123)
N	848		848	
Time Fixed effects	Yes		Yes	
Demeaned	Yes		Yes	

*Notes: Regressions used are bivariate PVAR(1) models using clustered standard errors. Both house prices and turnover-rates were instrumented using their second, third and fourth lag stated in levels using two stage GMM. Regressions are on the entire sample (2000-2013) and exclude outliers. ∆ indicates that the first difference is taken. Ln indicates that the natural logarithm is taken. * Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level.*

Table 2: Granger causality tests

	Chi-squared	p-value
(1) Turnover-rate→ House Price	0.267	0.605
(1) House Price→ Turnover-rate	4.652	0.031
(2) Turnover-rate→ Transaction Price	0.295	0.587
(2) House Price→ Turnover-rate	4.458	0.035

Notes: Numbers within parentheses correspond to the model number in Table 1. Turnover-rate→Transaction Price means that the turnover-rate Granger causes house prices. House Price→Turnover-rate means that house prices Granger cause the turnover-rates.

Urbanization model using bivariate PVAR(1) regressions

Similar to the panel ADL(1) model, the bivariate PVAR(1) model is run on urbanization level. The model is constructed in the same way as for the whole sample bivariate PVAR(1) model. Comparable to the entire sample panel ADL(1) model, variables are demeaned. The second lag of the mortgage interest rate, household income and the second lag of household income are added. Results can be found in Table 3.

Signs of coefficients are roughly in line with the panel ADL(1) model on urbanization levels shown in Table 12. However, this does not mean that there are no significant differences between both models. The magnitude of coefficients and even significance do shift significantly going from the panel ADL(1) model to the bivariate PVAR(1) model on urbanization level.

Just as in with the panel ADL(1) model on urbanization level, it is hard to say whether house price changes affect the number of sold homes more in densely populated areas than in thinly populated areas **(hypothesis 6)***.* The house price coefficient is only significant for moderate urban areas. Results are however in line with the panel ADL(1) model, where the coefficients were also mostly negative and insignificant.

Just as with the entire sample bivariate PVAR(1) model, Granger causality tests were run for the regressions. As can be seen from Table 4, there is no convincing evidence that the turnover-rate Granger causes house prices, or vice versa. Only house prices are significantly different from 0 in the regression for moderate-urban areas. This could be evidence that for these regressions that the bivariate PVAR specifications do not add much to the overall model, and we are safe to use the panel ADL(1) model instead.

Table 3: urbanization model (using a bivariate PVAR(1) regression)

	PVAR(1)		PVAR(1)		PVAR(1)	
	(3) Non-urban and low-urban		(4) Moderate urban		(5) High-urban and Very-high urban	
	∆(Turnover-Rate)	∆Ln(Transaction Price)	∆(Turnover-Rate)	∆Ln(Transaction Price)	∆(Turnover-Rate)	∆Ln(Transaction Price)
∆(Turnover-Rate)t-1	-0.085	-0.029	-0.306**	-0.134	0.048	-0.758
	(0.101)	(2.246)	(0.130)	(1.442)	(0.137)	(1.726)
∆Ln(House Price)t-1	-0.013	0.009	-0.025**	-0.180	-0.039	-0.639***
	(0.014)	(0.176)	(0.012)	(0.117)	(0.031)	(0.206)
∆(InterestRate)t-1	-0.199*	-1.92*	-0.299**	-0.028	-0.245	-5.210**
	(0.107)	(1.11)	(0.122)	(1.390)	(0.258)	(2.420)
∆(InterestRate)t-2	-0.303***	-2.56	-0.345**	-8.380***	-0.125	-2.480**
	(0.093)	(1.88)	(0.170)	(1.670)	(0.266)	(1.190)
∆Ln(Income)	0.003	0.086	-0.003	-0.571**	0.0250	0.612**
	(0.012)	(0.236)	(0.028)	(0.288)	(0.046)	(0.293)
∆Ln(Income)t-1	0.022**	0.085	-0.013	0.636***	-0.043	0.799**
	(0.011)	(0.209)	(0.022)	(0.182)	(0.039)	(0.352)
∆Ln(Income)t-2	-0.012	0.216	0.004	0.234	0.016	0.565
	(0.014)	(0.165)	(0.019)	(0.201)	(0.058)	(0.413)
N	440		240		240	
Time Fixed effects	Yes		Yes		Yes	
Demeaned	Yes		Yes		Yes	

*Notes: Regressions used bivariate PVAR(1) models with clustered standard errors. Both transaction prices and turnover-rates were instrumented using their second, third and fourth lag stated in levels using two stage GMM. Regressions are on the different urbanization levels indicated. The timespan is 2000-2013 and exclude outliers. ∆ indicates that the first difference is taken. Ln indicates that the natural logarithm is taken. * Indicates a significance at 10% level ** Indicates a significance at 5% level. *** Indicates a significance at 1% level.*

Table 4: Granger causality tests

	Chi-squared	p-value
(3) Turnover-rate→ House Price	0.000	0.990
(3) House Price→ Turnover-rate	0.907	0.341
(4) Turnover-rate→ House Price	4.537	0.033
(4) House Price→ Turnover-rate	0.009	0.926
(5) Turnover-rate→ House Price	0.193	0.661
(5) House Price→ Turnover-rate	1.548	0.213

Notes: Notes: Numbers within parentheses correspond to the model number in Table 3.

Turnover-rate→Transaction Price means that the turnover-rate Granger causes house prices.

House Price→Turnover-rate means that house prices Granger cause the turnover-rates.