Master's thesis

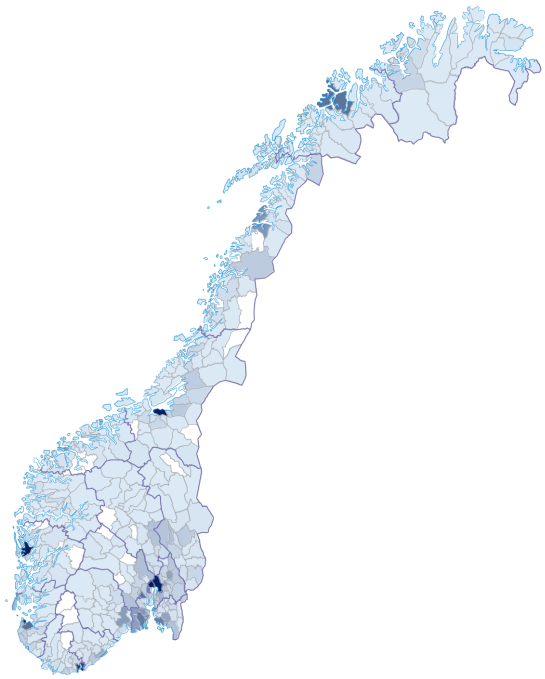
NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

Brekke, Sondre
Grov, Hans Martin
Pedersli, Trond Einar Moen

Local Population Characteristics and House Prices

Evidence From Norwegian Microdata

Master's thesis in Industrial Economics and Technology Management
Supervisor: Peter Molnár
June 2020



NTNU
Norwegian University of Science and Technology

Brekke, Sondre
Grov, Hans Martin
Pedersli, Trond Einar Moen

Local Population Characteristics and House Prices

Evidence From Norwegian Microdata

Master's thesis in Industrial Economics and Technology Management
Supervisor: Peter Molnár
June 2020

Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

Preface

This master's thesis is written by three students at the Department of Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU) in Trondheim during the spring of 2020. The thesis concludes a year of studying the Norwegian housing market, where we in a previous paper created and analysed several house price indices for selected municipalities. This thesis would not have been possible without help from friends, family and colleagues.

We would like to thank Alva Technologies for providing us with a detailed data set containing transactions of dwellings. Their data allowed us to dig deep into the Norwegian housing market.

We would also like to thank the Norwegian Centre for Research Data for giving us access to their database with yearly data on population and salary. Together with data collected from Statistics Norway on other explanatory variables, our final data set become more detailed than those used earlier.

Furthermore, we would like to thank the Norwegian Labour and Welfare Administration for their help with getting unemployment data from municipalities back to the early 1990s.

Finally, we would like thank our supervisor Peter Molnár for dedicating a lot of his time this last year to our research. He has provided us with great feedback and insight on financial research, and it has been a pleasure to work with him.

Abstract

This study combines socioeconomic measures and municipality characteristics from 381 municipalities in Norway with 1,204,506 actual sales of dwellings, to examine the relationship between nine variables and changes in house prices. This differs from existing literature which uses house price indices to represent the development in the housing market. We find that an increase in the population, the number of people that have completed higher education, the number of students relative to the population, the proportion of immigrants or the mean salary positively affect house prices. When further examining demographic factors, an increase in the proportion of the population aged over 66 is found to affect house prices negatively, while an increase in the proportion of the population aged between 18 and 66 is shown to have the opposite effect. Finally, an increase in income inequality or unemployment rate is found to have a negative effect.

Sammendrag

Dette studiet kombinerer sosioøkonomiske faktorer og kommunekarakteristikker fra 381 kommuner i Norge med 1.204.506 faktiske salgsdata på boliger, for å studere effektene av ni variabler på boligprisene. Dette skiller seg fra eksisterende litteratur som bruker boligprisindekser for å representere utviklingen i boligmarkedet. Vi finner at en økning i innbyggertall, antall personer som har fullført høyere utdanning, antall studenter relativt til folketall, andelen innvandrere eller gjennomsnittslønn påvirker boligprisene positivt. Ved nærmere demografiske undersøkelser kommer det fram at en økning i andelen over 66 år påvirker boligprisene negativt, mens en økning i andelen mellom 18 og 66 år har motsatt effekt. Videre viser resultatene at en økning i inntekstforskjell eller arbeidsledighet har en negativ effekt.

Abbreviations

B2C Business-to-Consumer
CEE Central and Eastern Europe
CV Coefficient of Variation
GDP Gross Domestic Product
HPI House Price Index
MSA Metropolitan Statistical Area
NAV The Norwegian Labour and Welfare Administration
NSD Norwegian Centre for Research Data
NUS The Norwegian Standard Classification of Education
OECD The Organisation for Economic Co-operation and Development
SSB Statistics Norway

Contents

List of Figures

List of Tables

1 Introduction

Purchasing a house is a big decision to make for most people. In addition to serving as a home, it also functions as a financial investment. Certain factors are commonly acknowledged to directly affect the housing market in different ways. Some examples of these are key figures reflecting the general level of a nation's economy such as the real GDP, the policy rate, stock market indices and the unemployment rate. Looking at other factors such as immigration, there is no consensus as to how, if at all, they affect house prices. Moreover, the increasing interdependence between nations worldwide, known as globalization, adds another level of complexity to the market dynamics. This became evident during the global financial crisis which affected housing markets all around the world. On a more local level, however, socioeconomic factors and characteristics that are specific to various regions are of high interest. Demographic measures such as income, population, age and educational level may vary dramatically across the different regions of a country. Hence, when studying the housing market of an economy, it is important to acknowledge the potential effects of these regional variations.

As of 2018, over 80% of all Norwegian households own their dwelling, which is relatively high when compared to neighbouring countries such as Sweden and Denmark, with 64.1% and 60.5% respectively (Eurostat, 2019). Therefore changes in the housing market are of high interest to most Norwegians. Also, high participation in the housing market in Norway makes it particularly relevant for research, because, relative to other countries, house prices are influenced more by the general population instead of just the high-income population. The market is furthermore considered very transparent, with most listings being published on the online marketplace FINN.no (Eiendom Norge, 2019) and 90% of all sales following the English auction standard (Olaussen et al., 2018). Moreover, several public institutions report detailed data on demographics and socioeconomic factors. Altogether, this makes Norway very well suited for studies on house prices. We benefit from these sources and create a large data set consisting of actual transactions of dwellings, instead of house price indices. Using this, we are studying the effects of changes in municipality characteristics on house prices.

Several papers discuss which factors that drive the house prices such as the variable mortgage rate (Otto, 2007; Tsatsaronis and Zhu, 2004), the stock market and real GDP (Pillaiyan, 2015), changes in migration, income and other demographics (Pashardes and Savva, 2009; Poterba et al., 1991). The effect of the different factors may vary across countries (Englund et al., 1996; Geng, 2018; Kasparova and White, 2001; Taltavull de La Paz and White, 2012). Case and Shiller (1990) use quarterly data from 1970-Q1 to 1987-Q3 for Atlanta, Chicago, Dallas and San Francisco to run a time series cross-sectional regression. The results state that both changes in per capita real income and in the adult population are positively related to price changes over the subsequent year. Anundsen and Jansen (2013) find that house prices depend on real disposable income, household borrowing and the housing stock in the long-run. Green and Hendershott (1996) measure how real house prices are impacted by demographics such as age structure, education and income. They find that holding all else constant, the ageing of the population does not seem to lower real house prices. Galati et al. (2011) investigate both macro and micro drivers of house price dynamics using Dutch data and find that household-specific and house-specific factors such as income and wealth, educational level, cohort and year of construction are strongly related to subjective house prices.

There are several studies on the effect of immigration on house prices. Saiz (2007) studies the local economic impact of immigration in U.S. destination cities and finds that average rent and housing values increase by about 1% if there is an immigration inflow equal to 1% of a city's population. Mussa et al. (2017) argue that an increase in immigration inflow in one particular MSA will also lead to increased rents and house prices in neighbouring MSAs. This positive relationship between immigration and house prices is also found in studies done in Switzerland and Spain (Degen and Fischer, 2009; Gonzalez and Ortega, 2013). Larkin et al. (2018) use 474 estimates on the impact of immigration on house prices from 14 developed countries. They find that, on average, immigration increases house prices. In countries that are less welcoming to immigrants, however, this increase is more limited. Sá (2015) draws a different conclusion, finding that immigration has a negative effect on the house prices in the UK. This is explained with the mobility response of the native population; a negative income effect arises when those at the top of the wage distribution leave areas with immigration, and thus these areas experience a decline in house prices. A study done by the central bank of

Norway concludes that when controlling for changes in the unemployment rate and domestic population, the house prices in an area will increase by about 3% when the proportion of immigrants increases with 1 percentage point of the original population (Nordbø, 2013).

Changes in the age structure of the population is another interesting topic when studying house prices. Mankiw and Weil (1989) study the effects of major demographic changes on the housing market in the U.S. They find that the increase in real house prices in the 1970s is mainly due to the baby boomer generation reaching its house-buying years. Levin et al. (2009) use a difference-in-difference methodology to examine how population ageing and population decline affect the house prices in Great Britain. The results show that both population decline and population ageing have negative impacts on the house prices. Using a data set with house prices from 22 advanced economies to investigate the effect of demographics on real house prices, Takáts (2012) also finds that an ageing population affects house prices negatively.

Inequality can be measured in many ways and is often a hot topic for debate in governments around the world. Goda et al. (2016) use both absolute and relative income inequality measures to test whether this affects house prices. They conclude that increasing income inequality contributed to the rise in real house prices, measured by yearly averages of HPIs, in the studied OECD countries between 1975-2010. Özmen et al. (2019), however, conclude that income inequality measured by the Gini coefficient is negatively correlated with house price changes, using regional data from Turkey. Määttänen and Terviö (2014) analyse six U.S. metropolitan areas, and conclude in similar terms, finding that increased income inequality on average has a modest but negative effect on house prices.

The unemployment rate is a measure which is frequently monitored by political leaders. When reported on a national level it is viewed as a key indicator of the health of the economy. Liu and Shen (2005) find that the housing prices of new dwellings in China decrease as the unemployment rate increases. Abelson et al. (2005) find that the long-run elasticity of real house prices is negative with respect to unemployment, using data from Australia.

Fack and Grenet (2010) investigate how the quality of education offered by schools nearby affect house prices using data from Paris. They find that a standard deviation increase in the performance of public schools, increase house prices by 1.4%-2.4%. Black (1999) examines

how parents value school quality measured through house prices. The results suggest that parents are willing to pay about 2.1% more for houses located in school districts where the test scores are 5% higher than the mean, i.e. approximately one standard deviation.

When conducting studies across various cities or countries, it is often convenient to use panel data. Gallin (2006) examines the long-run relationship between house prices and income by using a panel of 95 U.S. cities over 23 years. He finds that there is no evidence of cointegration between these variables. Goodhart and Hofmann (2008) use a panel of 17 industrialized countries over 37 years and find that there is evidence of a significant multidirectional link between house prices, monetary variables and the macroeconomy. Égert and Mihaljek (2007) study how eight transition economies of central and eastern Europe (CEE) and 19 OECD countries respond differently to changes in fundamental determinants of house prices, such as GDP per capita, real interest rates, housing credit and demographic factors. They find that, in addition to such conventional fundamentals, the house prices in CEE are also determined by institutional development of housing markets, and housing finance and quality effects. Mikhed and Zemčík (2009) use annual data from 1978 to 2007 on 22 MSAs to examine whether the U.S. housing market recently experienced a bubble. They conclude that there was such a house price bubble before 2006 and that the correction has not been sufficient.

The housing market is frequently discussed in Norway. Larsen and Sommervoll (2004) investigate what factors decide the prices in the Norwegian housing market, and describe a complex market driven by mortgage rates, changes in income and unemployment. They also point to the level of optimism or pessimism for the future among households and banks as an influential factor for the house prices. Moreover, urbanization, the increased number of single people and changes in the age distribution, typical family structure and immigration are said to affect the housing market. Jacobsen and Naug (2004) estimate an empirical model and aim to find the most important explanatory variables describing changes in Norwegian house prices. Their report conclude that newly constructed homes, income, interest rate and unemployment rate are the main factors. Larsen (2018) investigates how micro and macroeconomic factors affect the housing market in three of the largest cities in Norway, using two different empirical models; one of these being the model developed by Jacobsen and Naug (2004) and the second is a difference-in-difference model inspired by Levin et al. (2009). The study compares the results from the two methods and finds that unemployment

rate, interest rate, oil price and level of debt are statistically significant in all or some of the cities.

We contribute to the literature by studying the relationship between nine socioeconomic factors and changes in house prices. Unlike other studies, we use actual sales data of unique dwellings instead of house price indices. The data set contains more than 1.9 million sales from 1995 to 2015 from the whole of Norway, with nine variables describing characteristics across 381 municipalities. In 2015 these municipalities housed more than 98% of the Norwegian population. This allows us to study the housing market on a more precise level. Such a comprehensive data set has never been applied in the existing literature. Also, by including dummy variables for the municipalities and time periods in the panel, we can control for general economic development at municipality level. Hence, we can better explain the effects of these variables on house prices and describe both differences and similarities across the country. Therefore, the microeconomic depth of the data set makes this study an important extension of the existing literature.

First, Section 2 provides a detailed presentation of the data set. Section 3 describes the model, while the results are presented and discussed in Section 4. The findings are concluded in Section 5.

2 Data

The data set used in this study consists of data from four different sources. Section 2.1 elaborates on the data on sales of dwellings provided by Alva Technologies, while Section 2.2 presents the data on characteristics of Norwegian municipalities, provided by Microdata, NAV and SSB. Data cleaning is addressed in Section 2.3, and descriptive statistics of the data set are presented in Section 2.4.

2.1 Sales Data

Data on sales of dwellings for the whole of Norway is provided by Alva Technologies, a company developing an innovative and transparent solution to create a more efficient real estate market. They obtained the data from mainly two public registers; *Grunnboken*, containing all juridical binding transactions, and *Matrikkelen*, containing information about land and dwellings. These are both published by the Norwegian Mapping and Cadastre Authority. Alva Technologies have further improved this with data from both disclosed private and publicly accessible data sets, and user input from their B2C-service virdi.no. This service lets owners of dwellings correct information, as well as give an estimate of the quality of the dwelling. Altogether, this becomes a thorough data set covering the whole of Norway, with 1,904,532 sales from 1995 to 2015.

The data set contains two different prices per dwelling-transaction; the *sales price* and the *total price*. The former is the value of the bid that won the English auction. Most condominiums in Norway have common debt, especially newly built dwellings which sometimes have debt as large as the price of the dwelling. The *sales price* would therefore, in cases like this, be artificially low compared to the actual price of the dwelling. To account for this, the *total price* of the dwelling is used in our study (denoted *total_price* when referred to as a variable). This is also the price which SSB uses (Medby and Takle, 2020).

2.2 Municipality Characteristics

Yearly data on different characteristics and socioeconomic measures for each municipality are gathered and merged from three sources. Panel data on salary and number of people that have completed higher education are provided by Microdata, an online service funded by the Research Council of Norway and developed by SSB and NSD. Furthermore, SSB provides socioeconomic measures as population, distribution of age, number of immigrants and number of students. Finally, NAV issue monthly statistics on unemployment in Norway.

As several municipalities merged and/or changed their ID throughout the sample period, the data set required careful treatment to make sure the merge with the sales data was correct. The sales data from Alva Technologies use the municipality IDs from mid-2019, hence these IDs were used as a reference.

Thus, nine time-varying variables were created for each of the municipalities in Norway. These variables are shown in Table 1.

Variable	Source	Description
population	SSB	Number of inhabitants registered in the National Population Register.
population_middle	SSB	Number of inhabitants between the age of 18 and 66 in percentage of *population*.
population_old	SSB	Number of inhabitants over the age of 66 in percentage of *population*.
immigrants	SSB	Number of immigrants in percentage of *population*. Here, an immigrant is defined as a person born abroad of two foreign-born parents and four foreign-born grandparents.
higher_education	SSB	Number of inhabitants that have completed higher education (defined by NUS2000-level being equal to 6, 7 or 8) in percentage of *population*.
salary_CV	Microdata	Coefficient of variation for salary, given as the ratio of the standard deviation to the mean.
salary_mean	Microdata	Mean salary (including benefits in kind, sick pay and child benefits throughout the calendar year).
students	SSB	Number of students divided by the *population*. Here, the former consists of all the students registered at a university or college in that municipality, regardless of where they are registered in the National Population Register.
unemployment	NAV	Unemployed people in percentage of the workforce.

Table 1: Description of municipality characteristics used in this study.

2.3 Data Cleaning

For the data set to be in its most meaningful condition, thorough data cleaning is performed. On the data containing sales of dwellings, the following cleaning is done:

- Remove transactions where the *total_price* is missing or has a value of zero.
- If a dwelling has multiple sales within the same year, the average value of the *total_price* is used to represent the price for that year. As the municipality characteristics are recorded yearly, we require a dwelling to have a maximum of one *total_price* per year.
- If a dwelling is only sold once throughout the period, it is removed as we require *repeated sales*.
- Remove sales from municipalities where we do not have consistent data on either transactions or municipality characteristics throughout the sample period.

The following modifications are applied to the data on municipality characteristics:

- Microdata winsorize all numeric data on a 2% level. This means that the 1% highest values are set to the 99^{th} percentile, and the 1% lowest values are set to the 1^{st} percentile.
- The unemployment rate is reported monthly, thus the average of the 12 months are used to represent the year in our study. In some municipalities, however, a month or two are missing for a certain year. In these cases, we use a linear interpolation based on the closest preceding and subsequent months containing real data. If a municipality is missing data on more than two months during a year, the municipality is deleted to preserve the correctness of the data set. A similar linear interpolation is applied to the number of inhabitants with higher education, if data on one year is missing for a municipality.

2.4 Descriptive Statistics

After the cleaning procedure is complete and the data from the different sources are merged, we are left with a data set consisting of 1,204,506 sales from 381 of the 422 municipalities in Norway. These municipalities are housing more than 98% of the total population in Norway, based on 2015-numbers.

The dwellings in the data set are sold 2.56 times, on average. Figure 1 shows the distribution of sales during the sample period. 62.8% of the dwellings have only been sold twice and 24.4% have been sold three times, indicating that the average Norwegian family does not move very frequently.

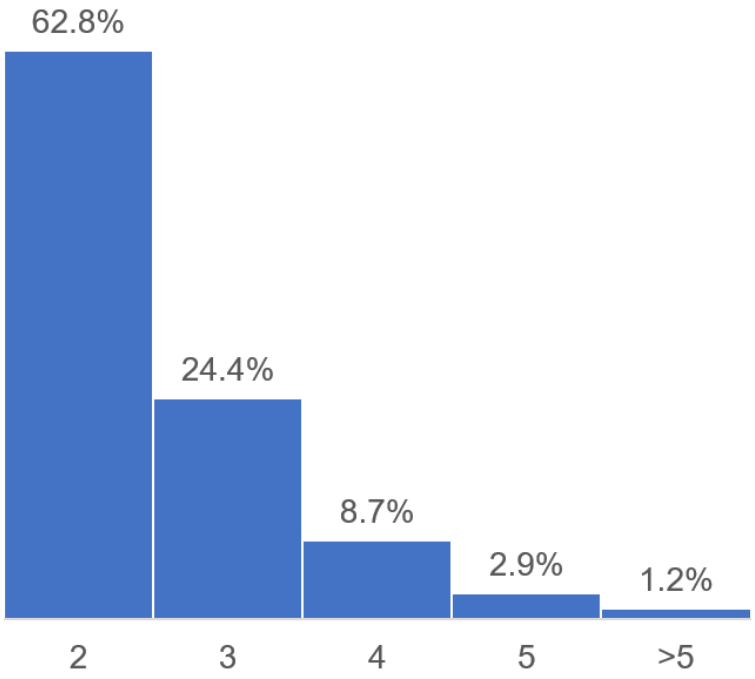


Fig. 1: Distribution of number of sales for unique dwellings in the data set.

Figure 2 displays the number of sales per year in the sample period. Sales of cooperative shares were not added to the public register *Grunnboken* before 2007 which explains the noticeable increase in the number of sales.

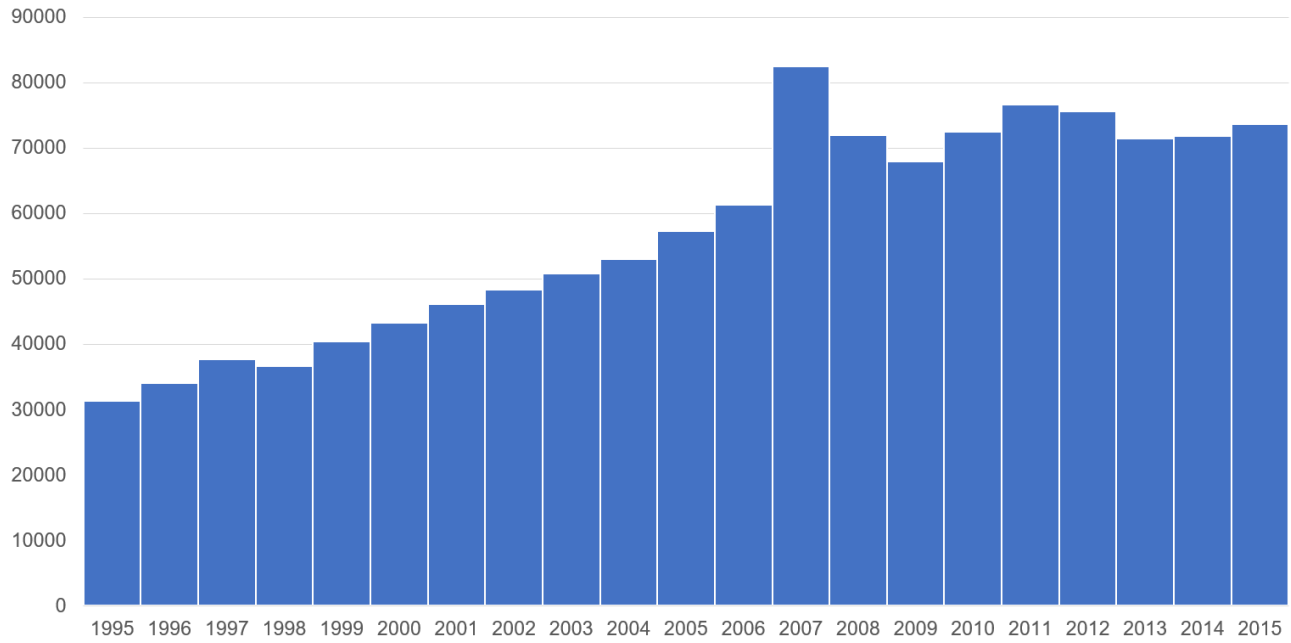


Fig. 2: Number of sales of dwellings for each year in the sample period.

The data set contains four different types of dwellings that are common in Norway:

1. *Apartment:* Several designated rooms to live in. Often part of a bigger complex.
2. *Detached:* Separate from other dwellings.
3. *Semi-detached:* Dwellings with one common wall shared with a similar dwelling.
4. *Terrace:* Similar dwellings in a row sharing multiple walls.

Figure 3 shows the proportion of these four types of dwellings in the data set. More than 70% of the transactions are sales of detached dwellings and apartments.

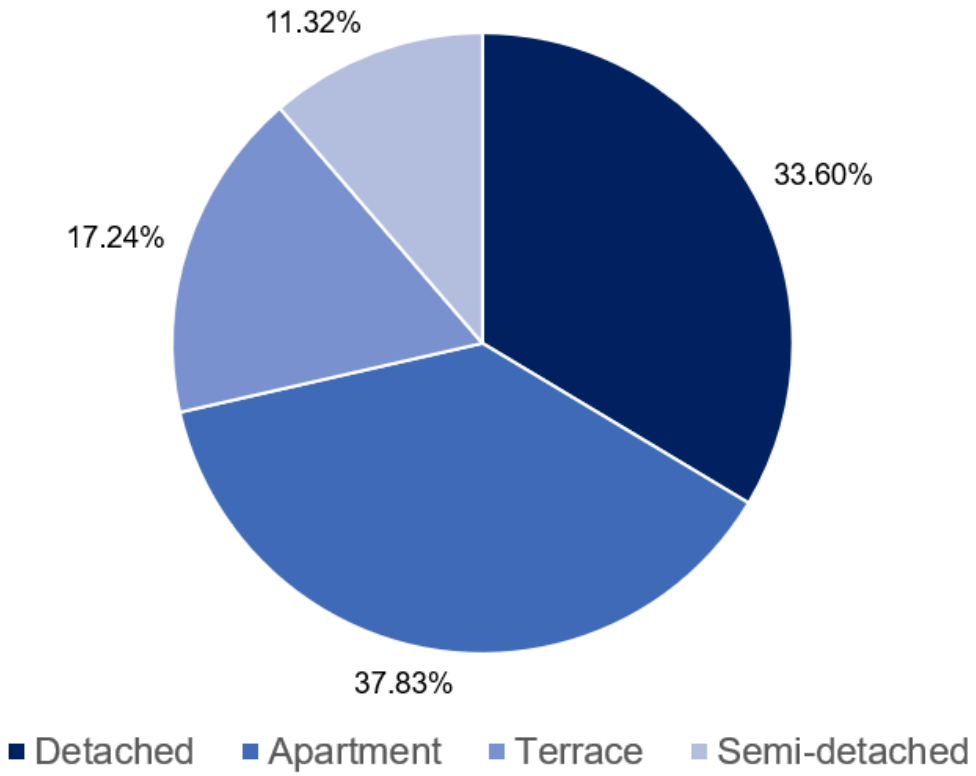


Fig. 3: Distribution of the various types of dwellings in the data set.

Figure 4 illustrates the number of years between sales of the same dwelling in the data set. The mean is 4.74 years, and the distribution is right-skewed. Interestingly, the mode is only 2 years, a short period considering the findings from Figure 1 and the belief that dwellings often are regarded as long-term investments.

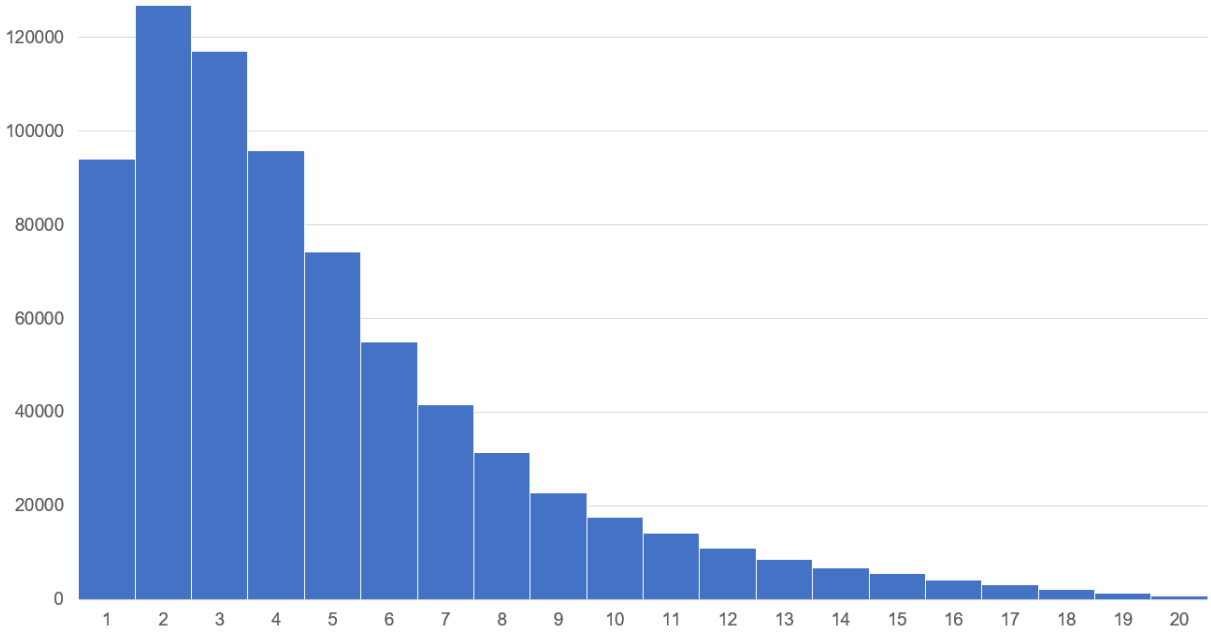


Fig. 4: Time (in years) between two sales of the same dwelling.

3 Methodology

In this section, we present the models and the theory behind them. A repeat sales approach is used, with the change in the price of a dwelling as the dependent variable. One major advantage of our detailed data set is that it allows us to use a simple model. In particular, a fixed effects model, with both time and municipality dummy variables, is applied when running panel data regressions.

3.1 Repeat Sales Pairs

The variables are divided into two sets (see Table 2). Variables in S^K are reported as absolute numbers, whereas those in S^Q are reported as a relative number.

Variables in S^K	**Variables in S^Q**
population	*higher_education*
salary_CV	*immigration*
salary_mean	*population_middle*
total_price	*population_old*
	students
	unemployment

Table 2: The variables are divided into two sets, S^K & S^Q.

Instead of using a complex hedonic method to account for all factors explaining the house price, we use the repeat sales method where the dwelling characteristics are assumed to be constant over time. The repeat sales method uses the change in house price between two sales of the same dwelling. Therefore, we construct *repeat sales pairs*, i.e. we arrange sales of the same dwelling in pairs. If a dwelling has been sold more than two times during the sample period, a non-overlapping decomposition is chosen when constructing repeat sales pairs. Hence, a dwelling sold in 2000, 2003 and 2008 would in the data set be presented as two repeat sales pairs describing consecutive price changes, i.e. 2000 $\rightarrow$ 2003 and 2003 $\rightarrow$ 2008. Thus, each line in the data set represents a single sales pair and contains the change in *total_price* as well as changes in the nine explanatory variables. Table 3 shows an example

of an entry in the data set. Note that the Address ID has been censored for privacy concerns. In total, the data set consists of 733,739 such sales pairs. For variables in S^K we take the logarithmic relative change, whereas for those in S^Q we take the simple difference. Hence, if $x \in S^K$ and $y \in S^Q$, then we use $\Delta \ln(x) = \ln(\frac{x_2}{x_1})$ and $\Delta y = y_2 - y_1$.

Entry	**Value**
Address ID	******
Municipality ID	403
Year of first sale	1997
Year of second sale	2003
Δ ln(*total_price*)	0.737599
Δ *higher_education*	1.967
Δ ln(*population*)	0.047621
Δ *immigrants*	1.582
Δ ln(*salary_mean*)	0.269378
Δ *population_middle*	0.193
Δ *population_old*	-1.385
Δ ln(*salary_CV*)	0.0
Δ *students*	11.44
Δ *unemployment*	0.017

Table 3: Example of a repeat sales pair from the data set.

Figure 5 shows a map of the 422 municipalities in Norway, where the municipalities are color-coded based on how many sales pairs that are included in the data set. As expected, the municipalities with the largest cities in Norway have the most sales pairs. Due to the geography of Norway, it is not surprising that the middle part of Norway is light blue, as this is mainly mountains.

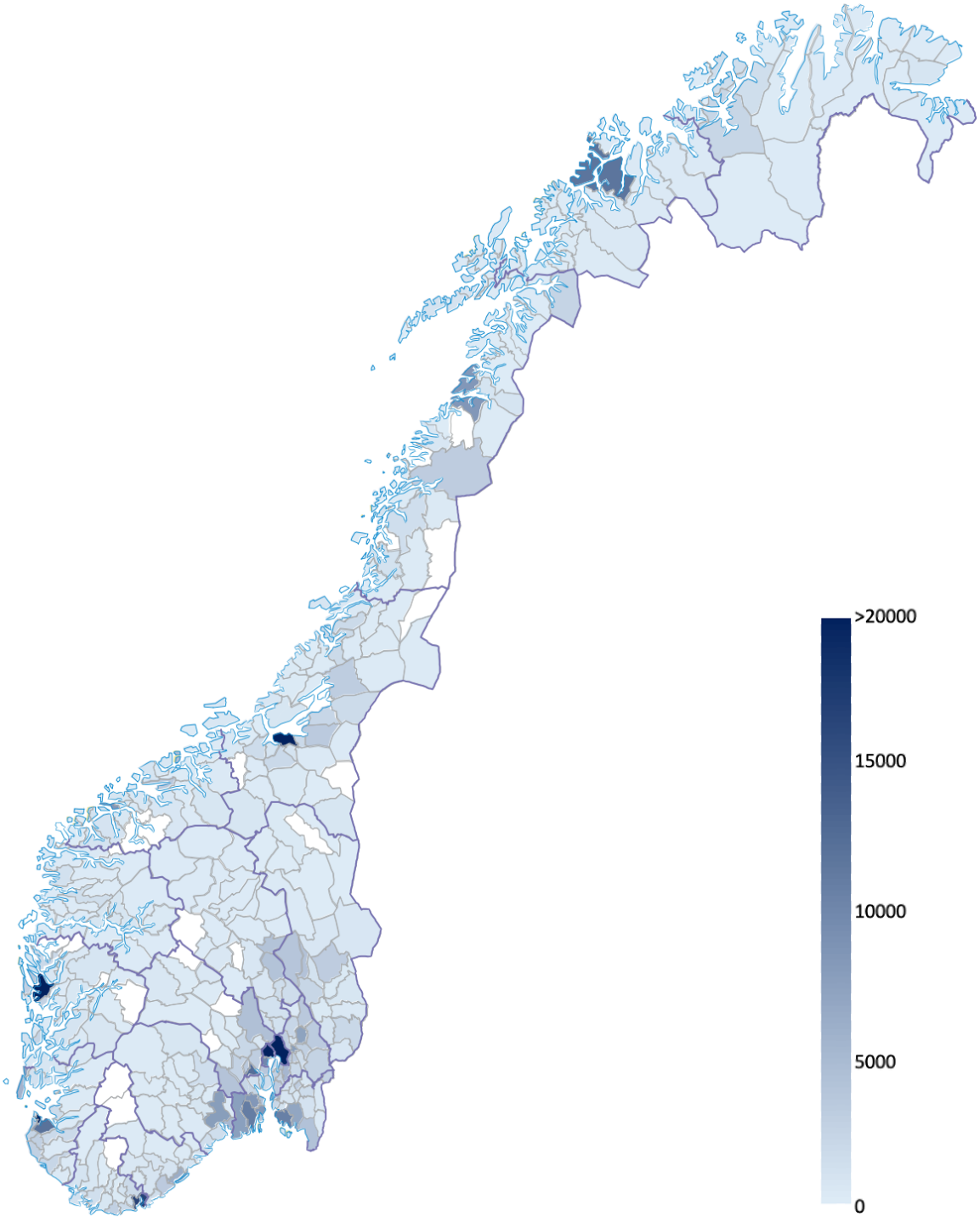


Fig. 5: Map of Norway showing the distribution of sales pairs across the municipalities. If a municipality is white, it means that either data is missing or there are no sales pairs during the sample period.

3.2 Model

A fixed effects approach is used for the regression model when analysing the historical sales data. To account for the geographical variations across Norway, municipality dummy variables are included. Furthermore, time dummy variables are also included to capture changes in the general level of the economy.

With the sets defined as in Table 2, the fixed effects model used in this study can be written as follows:

$$\ln(\frac{P_i^{mt'}}{P_i^{mt}}) = \alpha + \sum_{\tau=1}^{T} \delta^\tau D_i^\tau + \sum_{j=1}^{J} \mu^j D_i^j + \sum_{k \in S^K} b_k \cdot \ln(\frac{x_{ik}^{mt'}}{x_{ik}^{mt}}) + \sum_{q \in S^Q} c_q \cdot (y_{iq}^{mt'} - y_{iq}^{mt}) + \varepsilon_i^{mtt'} \quad (1)$$

where P_i^{mt}, is the *total price* of a dwelling i, located in municipality m, sold at time t ($t < t'$). Here α is the constant term and $b_1, ..., b_K$ are the price coefficients related to the explanatory variables $x_{i1}^{mt}, ..., x_{iK}^{mt}$ in set S^K. Similarly, $c_1, ..., c_Q$ are the price coefficients related to the explanatory variables $y_{i1}^{mt}, ..., y_{iq}^{mt}$ in set S^Q. The sample consists of $T + 1$ time periods (i.e. from year $t = 0$ to year $t = T$). D_i^τ is a time dummy variable, corresponding to the parameter δ^τ, with value 1 for all $t < \tau \leq t'$ and 0 otherwise. τ iterates from 1 to T, i.e. we skip the base period where $t = 0$ to avoid perfect collinearity. D_i^j is a municipality dummy variable, corresponding to the parameter μ^j, with value 1 if $j = m$ and 0 otherwise. Again we leave out one dummy variable (i.e. $J = 380$, one less than the total number of municipalities included in the data set). The error term, $\varepsilon_i^{mtt'}$, is assumed to have constant variance and expected value $\mathbb{E}[\varepsilon_i^{mtt'}] = 0$.

In addition to running (1), a regression without any dummy variables is run so that the consequence of including fixed effects become evident. Section 4 also includes various models where some of the variables are left out due to being highly correlated with each other. Also, results from univariate regressions with and without fixed effects are included.

3.3 Dealing With Autocorrelation and Heteroscedasticity

If the error terms from the regression are heteroscedastic, the variance changes over time and violate the underlying assumption of homoscedasticity, i.e. $var(\varepsilon_t) = \sigma^2 < \infty$. A method of dealing with heteroscedasticity when the form is unknown, is to use heteroscedasticity robust standard errors or White-Huber-Eicker robust standard errors. This paper uses the former.

Another basic assumption when running least squares regressions is that the error terms are uncorrelated, i.e. $Cov(\varepsilon_i, \varepsilon_j) = 0$ $(i \neq j)$. If this is not the case, they are said to be autocorrelated. If a dwelling is sold n times, it will generate $n - 1$ sales pairs. Still, one particular sale will only be part of a maximum of two sales pairs. Therefore, if a sales price is incorrect, it will only affect a maximum of two sales pairs. Hence, the model will not have a problem with autocorrelation.

4 Results and Discussion

This section will present and discuss the findings from this study. First, a correlation matrix showing how the changes in variables are correlated with each other is presented in Figure 6. Then, the results from the panel data regressions for five different models are shown in Table 4. Furthermore, the results from the univariate regressions both with and without fixed effects, are presented in Table 5. Finally, the findings are discussed and compared to those from existing literature.

Figure 6 presents how the changes in the variables, i.e. the regression inputs listed in Table 3, correlate with each other. It suggests that $\Delta \ln(salary\_mean)$, $\Delta\ higher\_education$, $\Delta\ immigrants$ and $\Delta \ln(population)$ are highly positively correlated with each other, all with correlation values above 0.88. Due to a concern that this potentially could affect the results, we choose to include four additional models. Each of these models include only one of these four variables and are presented in Table 4.

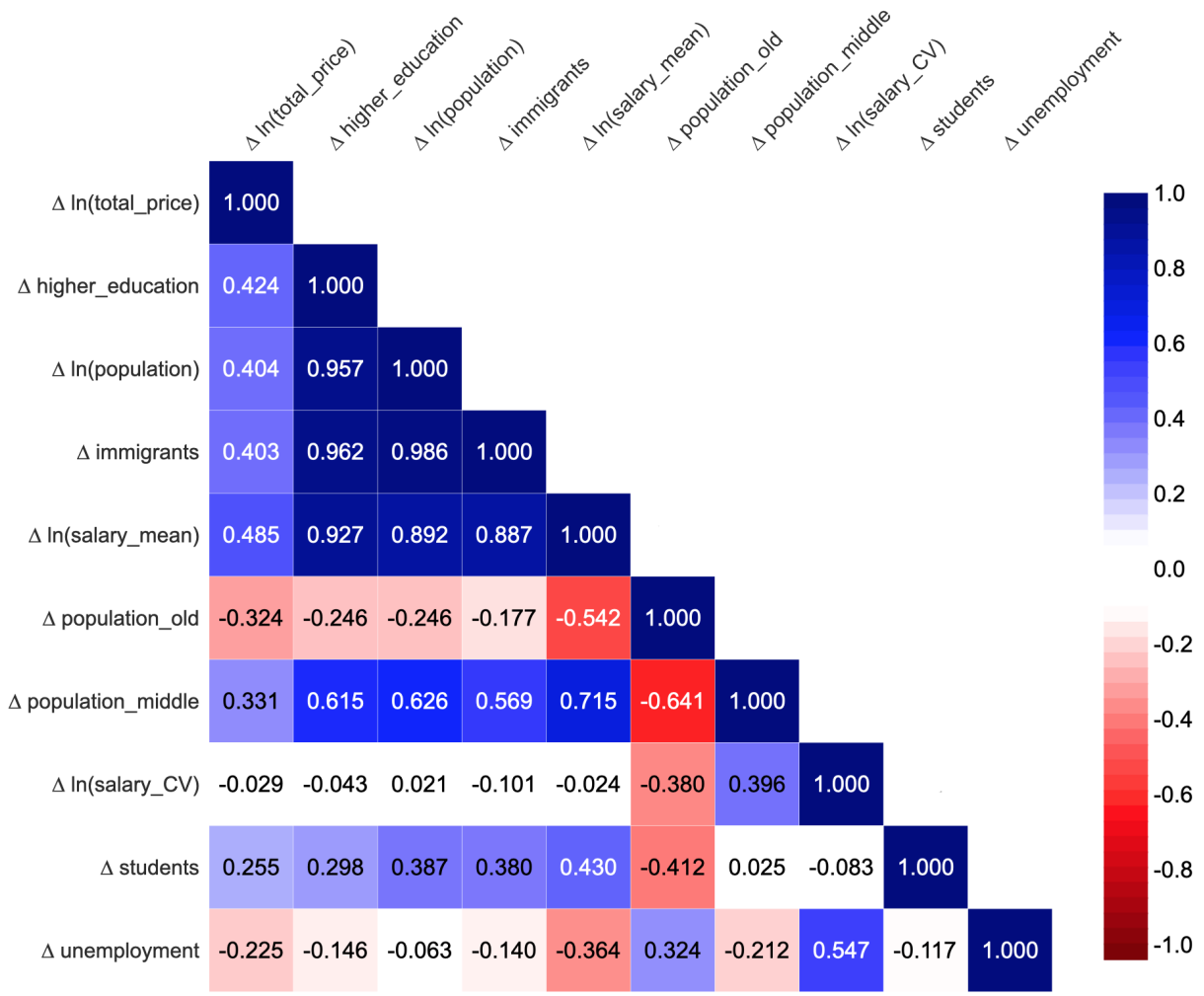


Fig. 6: Correlation matrix showing the correlation between changes in the variables under study.

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
alpha	0.0618**	0.0618**	0.0618**	0.0618**	0.0618**
	(0.0079)	(0.0079)	(0.0079)	(0.0079)	(0.0079)
Δ *higher_education*	0.0930**	-	-	-	0.0709**
	(0.0005)	-	-	-	(0.0008)
Δ ln(*population*)	-	0.0083**	-	-	0.0014**
	-	(0.0001)	-	-	(0.0001)
Δ *immigrants*	-	-	0.1109**	-	0.0504**
	-	-	(0.0007)	-	(0.0011)
Δ ln(*salary_mean*)	-	-	-	0.0476**	0.0127**
	-	-	-	(0.0002)	(0.0001)
Δ *population_old*	-0.0475**	-0.0333**	-0.0422**	-0.0327**	-0.0480**
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Δ *population_middle*	0.0225**	0.0512**	0.0243**	0.0496**	0.0167**
	(0.0008)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Δ ln(*salary_CV*)	-0.0011**	-0.0027**	-0.0014**	-0.0026**	-0.0009**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Δ *students*	0.0049**	0.0080**	0.0036**	0.0078**	0.0036**
	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Δ *unemployment*	-0.0461**	-0.0518**	-0.0446**	-0.0508**	-0.0439**
	(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0006)
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	733,739	733,739	733,739	733,739	733,739
R-squared	0.26	0.26	0.26	0.26	0.26

Table 4: Results of running five models with Δ ln(*total_price*) as the dependent variable. Model (1) - Model (4) are described in the start of Section 4, while Model (5) is described in Section 3.2. Heteroscedasticity robust standard errors are included in parentheses for all models. * and ** denote statistical significance at the 1% and 0.1% levels, respectively.

Using the data set described in Section 2.4, the results in Table 4 show that all of the variables are significant at the 0.1% level when accounting for general economic development over time and geographic variations across municipalities. This furthermore underlines the strength of the comprehensive data set.

The first explanatory variable, Δ *higher_education*, is shown to have a significant *positive* effect on changes in house prices, even when including $\Delta \ln(salary\_mean)$. Hence, this indicates that Δ *higher_education* explains changes in house prices beyond the effects of $\Delta \ln(salary\_mean)$. A potential explanation for this is that people with higher education may be more aware of the potential returns from investing in real estate. Hence, they may invest a larger percentage of their savings in the housing market. Furthermore, they are likely to have a bigger expectation of increased income in the future compared to their counterparts.

The results suggest that $\Delta \ln(population)$ has a significant *positive* effect on $\Delta \ln(total\_price)$. In general, an increase in population is likely to increase the demand for houses, and thus positively affect house prices using simple economic theory. Furthermore, the results indicate that Δ *immigrants* has a significant *positive* effect on $\Delta \ln(total\_price)$. Apart from immigrants, the population growth in a municipality is composed of net inflow of non-immigrants and a possible birth surplus. In the short run, one would think that the demand for dwellings is not heavily affected by births. On the other hand, as most immigrants are between the age of 18 and 59 (Statistics Norway, 2019a), they are likely to increase the demand right away. This reasoning could explain why the house prices will be positively affected if there is an increase in the proportion of immigrants in the population.

The results reveal some interesting findings regarding the changes in the age structure of the population. Δ *population_old* is found to have a significant *negative* effect on $\Delta \ln(total\_price)$. Hence, an increase in the proportion of the population aged over 66 has a negative effect on changes in house prices. On the other hand, an increase in the proportion of the population between the age of 18 and 66 will *positively* affect house prices. This is not surprising, as one would expect that the age group with the highest demand for dwellings is somewhere between the age of 18 and 66.

As an increase in $\Delta \ln(salary\_mean)$ leads to an increase in average household purchasing power, it is not surprising that this variable has a significant *positive* effect on $\Delta \ln(total\_price)$. On the other hand, an increase in Δ *unemployment* will affect $\Delta \ln(total\_price)$ *negatively*. Unemployed people may be worried about the future and concerned with precautionary saving rather than investing in a dwelling. Furthermore, an increase in salary inequality has a *negative* effect on $\Delta \ln(total\_price)$. This might be due to an increase in inequality causing less pressure on the medium-priced dwellings, and thus spreading the demand for houses more evenly over the whole distribution.

Finally, Δ *students* has a significant *positive* effect on $\Delta \ln(total\_price)$. One explanation for this positive effect is that an increase in students is likely to increase the actual number of inhabitants, as not all students necessarily are accounted for in the *population*-variable, as explained in Table 1. The reason for this is that not all students register that they have moved to a different city in the National Population Register.

Table 5 shows the results of running univariate regressions and various models with only fixed effects, to further investigate what drives the R-squared measure.

	Model (6)	Model (7)	Model (8)	Model (9)	Model (10)	Model (11)	Model (12)	Model (13)	Model (14)	Model (15)	Model (16)	Model (17)	Model (18)
alpha	0.0618**	0.0599**	0.3889**	0.0793**	0.0989**	0.0990**	0.0485**	0.3088**	0.2393**	0.3832**	0.3086**	0.3582**	0.0605**
	(0.0079)	(0.0009)	(0.0086)	(0.0010)	(0.0010)	(0.0010)	(0.0009)	(0.0006)	(0.0007)	(0.0006)	(0.0007)	(0.0006)	(0.0079)
Δ *higher_education*	-	-	-	0.1430**	-	-	-	-	-	-	-	-	0.0853**
	-	-	-	(0.0004)	-	-	-	-	-	-	-	-	(0.0027)
Δ ln(*population*)	-	-	-	-	7.1977**	-	-	-	-	-	-	-	-1.1479**
	-	-	-	-	(0.0214)	-	-	-	-	-	-	-	(0.1736)
Δ *immigrants*	-	-	-	-	-	0.1963**	-	-	-	-	-	-	0.0807**
	-	-	-	-	-	(0.0006)	-	-	-	-	-	-	(0.0050)
Δ ln(*salary_mean*)	-	-	-	-	-	-	1.5987**	-	-	-	-	-	0.2222**
	-	-	-	-	-	-	(0.0038)	-	-	-	-	-	(0.0308)
Δ *population_old*	-	-	-	-	-	-	-	-0.1986**	-	-	-	-	-0.1643**
	-	-	-	-	-	-	-	(0.0007)	-	-	-	-	(0.0024)
Δ *population_middle*	-	-	-	-	-	-	-	-	0.2577**	-	-	-	-0.0816**
	-	-	-	-	-	-	-	-	(0.0009)	-	-	-	(0.0022)
Δ ln(*salary_CV*)	-	-	-	-	-	-	-	-	-	-0.9421**	-	-	-0.7402**
	-	-	-	-	-	-	-	-	-	(0.0385)	-	-	(0.0690)
Δ *students*	-	-	-	-	-	-	-	-	-	-	0.0315**	-	-0.0007*
	-	-	-	-	-	-	-	-	-	-	(0.0002)	-	(0.0003)
Δ *unemployment*	-	-	-	-	-	-	-	-	-	-	-	-0.1067**	-0.0319**
	-	-	-	-	-	-	-	-	-	-	-	(0.0005)	(0.0012)
Time fixed effects	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No
Municipality fixed effects	Yes	No	Yes	No	No	No	No	No	No	No	No	No	No
Number of observations	733,739	733,739	733,739	733,739	733,739	733,739	733,739	733,739	733,739	733,739	733,739	733,739	733,739
R-squared	0.26	0.25	0.006	0.18	0.16	0.16	0.24	0.11	0.11	0.0009	0.06	0.05	0.25

Table 5: Results from univariate regressions and models with only fixed effects, using Δ ln(*total_price*) as the dependent variable. Heteroscedasticity robust standard errors are included in parentheses for all models. * and ** denote statistical significance at the 1% and 0.1% levels, respectively.

Looking at Model (6), the R-squared value is identical to those of the models in Table 4. This suggests that the variables in this study do not explain the variations in house prices beyond what is explained by the fixed effects. This is in line with what one would expect, as we have left out important macroeconomic factors. Rather, we aim to investigate the effects of specific municipality characteristics on house prices.

The results from running Model (7) reveal that the time dummy variables explain roughly 25% of the variations in house prices. The time dummy variables account for the general economic growth, as well as other macroeconomic variables that change over time, like the policy rate and mortgage regulations. On the other hand, the R-squared value of Model (8) which only includes municipality fixed effects, is only slightly greater than zero. This is because the explanatory variables in this case, i.e. the municipality dummy variables, are purely cross-sectional. Moreover, price levels may vary across municipalities, while the *relative price changes* are likely to be more similar. Finally, Model (9) - (17) show the results of running univariate regressions. This allows us to better understand how the different variables explain the changes in house prices on their own. As expected, $\Delta \ln(salary\_mean)$ gives the highest value of R-squared.

The existing literature is not directly comparable to this study, as we use a much larger data set with cross-sectional data on municipalities, in addition to actual transactions of dwellings instead of house price indices. Therefore, this study can be seen as an improvement to the existing work rather than a similar study, as using actual sales data allows for a more comprehensive analysis. However, all of the following papers have one or more similarities to this study, and it is therefore relevant to look at how the results of these compare to our findings.

Aligned with the results from our study, Levin et al. (2009) find that population ageing and population decline are pressing the house prices downward, using data from England and Scotland. However, they are using the average real house price instead of actual sales data. The same effect of ageing is found by Takáts (2012) who studies house prices in 22 advanced economies between 1970 and 2009, but again this relies on the use of house price indices instead of sales data.

The effects of immigration on house prices is especially interesting as there is no clear con-

sensus in the literature. Sá (2015) uses local house price indices in the UK and finds that immigration has a negative effect which is said to be due to the mobility response of the top-earning population. Nordbø (2013) is a more comparable study, as it accounts for the effects of several other variables like salary-growth and change in unemployment rate using house price statistics from Norway. His findings are in line with our results, suggesting that an increase in the proportion of immigrants positively affects house prices.

The impact of proximity to high-scoring schools on house prices are previously studied in the literature. However, the effect of changes in the students to population ratio is not studied before in this context. Moreover, an increase the number of students in a municipality could also be a sign of improved schools in this municipality, and as Black (1999) finds, prices are higher in areas where the schools perform better.

Jacobsen and Naug (2004) use a monthly price index for used housing to study the Norwegian market. In line with our findings, they conclude that the unemployment rate is significant *negative*. They also find that an increase in total wage income in the economy affects house prices *positively*.

5 Conclusion

This study aimed to examine the relationship between nine socioeconomic variables and changes in house prices. Unlike existing studies on this topic, we do not use a house price index or a median index to track the changes in house prices. Instead, we use actual transaction data on sales of dwellings in Norway from 1995 to 2015. In total, 1,204,506 sales from 381 municipalities in Norway, where more than 98% of the total population are registered (based on 2015-numbers), were included. We contribute to the literature with our approach of using actual transactions to examine the effects of various factors on changes in house prices, giving detailed and highly interesting results.

We conclude that several socioeconomic factors are found to have a significant effect on the changes in house prices. Our findings agree with the existing literature, that increases in population or mean salary have positive effects on the house prices. We also find that an increase in the ratio of students to population, the number of inhabitants that have completed higher education and the proportion of the population between 18 and 66 in a municipality, all positively affect the house prices. Finally, in contrast to some other studies, we find that an increase in the proportion of immigrants in a municipality positively affects house prices.

On the other hand, this study concludes that some variables affect house prices negatively. In line with the existing literature, an increase in the unemployment rate is shown to have a negative effect on house prices. In addition, we also find that an increase in the proportion of people over the age of 66 has a negative effect on house prices. Finally, using the coefficient of variation for salary, we find that an increase in income inequality affects house prices negatively.

Further research can be done by including more municipality characteristics and socioeconomic measures, which could give an even better understanding of the dynamics affecting the house pricing market. However, a major challenge is to obtain such data for all municipalities over the sample period. For instance, Gini-coefficients were not available on a municipality level. Another variable we tried to include was property tax, and again this proved hard to find as there is no national register and it can even vary across different areas within a municipality. Also, the inclusion of crime rates could be interesting, but these data were not publicly available for the whole sample period.

Using the results of this study and future findings, it could be interesting to create improved local pricing models. The main benefit of such an automated valuation model is to give more precise price estimates than brokers currently can offer. Furthermore, one could use machine learning models with both macro and micro factors to create accurate prediction models for house prices.

6 References

Abelson, P., Joyeux, R., Milunovich, G., Chung, D., 2005. Explaining House Prices in Australia: 1970–2003. Economic Record 81, S96–S103.

Anundsen, A.K., Jansen, E.S., 2013. Self-reinforcing Effects Between Housing Prices and Credit. Journal of Housing Economics 22, 192–212.

Black, S.E., 1999. Do Better Schools Matter? Parental Valuation of Elementary Education. The Quarterly Journal of Economics 114, 577–599.

Case, K.E., Shiller, R.J., 1990. Forecasting Prices and Excess Returns in the Housing Market. Real Estate Economics 18, 253–273.

Degen, K., Fischer, A.M., 2009. Immigration and Swiss House Prices .

Égert, B., Mihaljek, D., 2007. Determinants of House Prices in Central and Eastern Europe. Comparative Economic Studies 49, 367–388.

Eiendom Norge, 2019. Eiendom Norges boligprisstatistikk. Web site. Eiendom Norge. URL: https://eiendomnorge.no/boligprisstatistikk/.

Englund, P., Ioannides, Y.M., et al., 1996. House Price Dynamics: An International Empirical Perspective. Department of Economics, Uppsala University.

Eurostat, 2019. Distribution of Population by Tenure status, Type of Household and Income Group - EU-SILC Survey. Data table. Eurostat. URL: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_lvho02&lang=en.

Fack, G., Grenet, J., 2010. When do Better Schools Raise Housing Prices? Evidence from Paris Public and Private Schools. Journal of Public Economics 94, 59–77.

Galati, G., Teppa, F., Alessie, R.J., 2011. Macro and Micro Drivers of House Price Dynamics: An Application to Dutch Data. Available at SSRN 1803087 .

Gallin, J., 2006. The Long-run Relationship Between House Prices and Income: Evidence from Local Housing Markets. Real Estate Economics 34, 417–438.

Geng, M.N., 2018. Fundamental Drivers of House Prices in Advanced Economies. International Monetary Fund.

Goda, T., Stewart, C., Torres, A., 2016. Absolute Income Inequality and Rising House Prices. Center for Research in Economics and Finance (CIEF), Working Papers .

Gonzalez, L., Ortega, F., 2013. Immigration and Housing Booms: Evidence from Spain. Journal of Regional Science 53, 37–59.

Goodhart, C., Hofmann, B., 2008. House Prices, Money, Credit, and the Macroeconomy. Oxford Review of Economic Policy 24, 180–205.

Green, R., Hendershott, P.H., 1996. Age, Housing Demand, and Real House Prices. Regional Science and Urban Economics 26, 465–480.

Jacobsen, D.H., Naug, B.E., 2004. Hva driver boligprisene. Penger og kreditt 4, 229–240.

Kasparova, D., White, M., 2001. The Responsiveness of House Prices to Macroeconomic Forces: A Cross-Country Comparison. European Journal of Housing Policy 1, 385–416.

Taltavull de La Paz, P., White, M., 2012. Fundamental Drivers of House Price Change: The Role of Money, Mortgages, and Migration in Spain and the United Kingdom. Journal of Property Research 29, 341–367.

Larkin, M., Doucouliagos, H., Dubelaar, C., Klona, M., Newton, J., Stanley, T., Vocino, A., et al., 2018. Do House Prices Sink or Ride the Wave of Immigration? .

Larsen, B., 2018. An Analysis on the Norwegian Housing Market: How Can We Explain Recent Changes in the Housing Prices in Norway? A Regional Approach. Master's thesis.

Larsen, E.R., Sommervoll, D.E., 2004. Hva bestemmer boligprisene? .

Levin, E., Montagnoli, A., Wright, R.E., 2009. Demographic Change and the Housing Market: Evidence from a Comparison of Scotland and England. Urban Studies 46, 27–43.

Liu, H., Shen, Y., 2005. Housing Prices and General Economic Conditions: An Analysis of Chinese New Dwelling Market. Tsinghua Science and Technology 10, 334–343.

Määttänen, N., Terviö, M., 2014. Income Distribution and Housing Prices: An Assignment Model Approach. Journal of Economic Theory 151, 381–410.

Mankiw, N.G., Weil, D.N., 1989. The Baby Boom, the Baby Bust, and the Housing Market. Regional Science and Urban Economics 19, 235–258.

Medby, P., Takle, M., 2020. Price Index for Existing Dwellings. URL: https://www.ssb.no/en/bpi/.

Mikhed, V., Zemčík, P., 2009. Do House Prices Reflect Fundamentals? Aggregate and Panel Data Evidence. Journal of Housing Economics 18, 140–149.

Mussa, A., Nwaogu, U.G., Pozo, S., 2017. Immigration and Housing: A Spatial Econometric Analysis. Journal of Housing Economics 35, 13–25.

Nordbø, E.W., 2013. Innvandring og boligpriser i Norge .

Norwegian Centre for Research Data, 2015. Lønnsinntekter. Data table. URL: https://microdata.no/discovery/variable/?revision=1&name=INNTEKT_WLONN&date=1993-01-01&language=en.

Norwegian Labour and Welfare Administration, 2019. Historisk statistikk om arbeidsmarkedet - Kommune. Beholdning måned. 1995-2019. Data table. URL: https://www.nav.no/no/nav-og-samfunn/statistikk/arbeidssokere-og-stillinger-statistikk/historisk-statistikk.

Olaussen, J.O., Oust, A., Sønstebø, O.J., 2018. Bidding Behavior in the Housing Market under Different Market Regimes. Journal of Risk and Financial Management 11, 41.

Otto, G., 2007. The Growth of House Prices in Australian Capital Cities: What do Economic Fundamentals Explain? Australian Economic Review 40, 225–238.

Özmen, M.U., Kalafatcılar, M.K., Yılmaz, E., 2019. The Impact of Income Distribution on House Prices. Central Bank Review 19, 45–58.

Pashardes, P., Savva, C., 2009. Factors Affecting House Prices in Cyprus: 1988-2008. Econ. Policy Rev. 3.

Pillaiyan, S., 2015. Macroeconomic Drivers of House Prices in Malaysia. Canadian Social Science 11, 119–130.

Poterba, J.M., Weil, D.N., Shiller, R., 1991. House Price Dynamics: The Role of Tax Policy and Demography. Brookings Papers on Economic Activity 1991, 143–203.

Sá, F., 2015. Immigration and House Prices in the UK. The Economic Journal 125, 1393–1424.

Saiz, A., 2007. Immigration and Housing Rents in American Cities. Journal of Urban Economics 61, 345–371.

Statistics Norway, 2018. Utdanningsnivå, etter kommune og kjønn (K) 1970 - 2018. Data table. URL: https://www.ssb.no/statbank/table/09429.

Statistics Norway, 2019a. Innvandrere etter innvandringsgrunn. Data table. URL: https://www.ssb.no/statbank/table/06313/.

Statistics Norway, 2019b. Studenter i høyere utdanning i Norge. Kjønn og skolekommune (K) 2000 - 2019. Data table. URL: https://www.ssb.no/statbank/table/03814.

Statistics Norway, 2020a. Alders- og kjønnsfordeling i kommuner, fylker og hele landets befolkning (K) 1986 - 2020. Data table. URL: https://www.ssb.no/statbank/table/07459.

Statistics Norway, 2020b. Innvandrere og norskfødte med innvandrerforeldre. Data table. URL: https://www.ssb.no/statbank/table/07110.

Takáts, E., 2012. Aging and House Prices. Journal of Housing Economics 21, 131–141.

Tsatsaronis, K., Zhu, H., 2004. What Drives Housing Price Dynamics: Cross-Country Evidence. BIS Quarterly Review, March .

NTNU

Norwegian University of
Science and Technology