Capitalization of neighbourhood diversity and segregation

Viggo Nordvik
NOVA, Oslo and Akershus University College of Applied Sciences, Norway
Liv Osland
Western Norway University of Applied Sciences, Norway
Inge Thorsen
Western Norway University of Applied Sciences, Norway
Ingrid Sandvig Thorsen
University of Bergen, Norway

Abstract

In this paper, we study how concentrations and diversity of different groups of households were reflected in the housing prices of neighbourhoods in the Oslo urban area, Norway. The focus is primarily on the settlement pattern of immigrants, but the analysis controls for socio-economic and demographic characteristics. Based on a hedonic conditional autoregressive spatial model formulation, we find that households on average prefer neighbourhoods with a high concentration of natives, many immigrants from Western countries, and, at the same time, a diverse, thin representation of neighbours from a wide range of countries. We do not find that immigrants from specific countries or continents have a substantial negative impact on housing prices in a neighbourhood.

JEL classification: J15, R21, R23

1. Introduction

Over the last 30 years, many European cities have experienced a large increase in the population with non-native country background. This is the case also for Oslo, the capital of Norway. It is a challenging empirical question to study whether and how the changed population composition and its uneven distribution over the urban landscape affect levels and spatial patterns of house prices. In part, the challenges are due to the dual roles that house prices play, both as attractivity measures of the combined package of housing and locational characteristics, and as important constraints on the residential opportunity sets for immigrants and natives. It is also challenging because of the multidimensionality of the sets of characteristics that feed into house price variation, and the fact that many of these

This is a peer reviewed, accepted postprint-version of the following journal article: Nordvik V, Osland L, Thorsen I, Thorsen IS. Capitalization of neighbourhood diversity and segregation. Environment and Planning A: Economy and Space. 2019; 51 (8): 1775-1799. DOI: https://doi.org/10.1177/0308518X19861108

characteristics correlate. In particular, this is the case for neighbourhood compositional characteristics, which are the focus of this study.

We address these challenges in an empirical analysis of spatial home price variation in the Oslo region. The purpose of our paper is to model empirically both concentrations and diversity of different groups of immigrants, and to study how these spatial patterns are reflected in local housing prices. Our analyses complement prior work in three important ways. First, we account for local spatial interdependencies in a rich and flexible way using a spatial conditional autoregressive (CAR) modelling framework, which includes spatially correlated random effects. Osland et al. (2016) have demonstrated that this modelling approach largely adjusts for missing information on spatial characteristics, and contributes considerably to more accurate predictions of house prices. This is particularly so for heterogeneous urban areas, where controlling for omitted spatial variables is highly important. Second, we utilize a detailed and rich register-based data set to describe both country background and socio-economic composition of neighbourhoods. These two features of our approach strongly reduce the problems of potential bias due to omitted variables. Finally, we introduce and test a measure of diversity in neighbourhoods in terms of country background of inhabitants. In brief, this flexible measure counts the number of country backgrounds that are significantly present in a neighbourhood. Using this approach, and by analysing patterns in Oslo, we aim to contribute to the knowledge on the still unsettled question of how increased country background diversity affects spatial residential patterns in European cities.

Systematic empirical exploration of this issue is under-researched (Ellen and O'Regan, 2010) and the economic literature on the integration of immigrants has mainly focused on labour-market processes (Saiz and Wachter, 2011). As emphasized in the OECD/European Union report (2015), integration is furthermore an important social and spatial process. Hence, the physical presence, residential location, and social interactions of immigrants and natives are also major concerns in this respect (Vang, 2010). A predominant concentration and segregation of immigrants in, for instance, deprived neighbourhoods may counteract effective integration, and result in unfavourable individual outcomes, such as reduced social mobility, poorer access to work and deterioration of health. Therefore, it is important to understand the various drivers of physical separation or integration of immigrants in the spatial residential landscape. One such driver is the structure of housing markets, and the implicit price structure. Our analysis is based on the premise that the price structure together with other characteristics of housing markets is fundamental for the allocation of people and groups of people at different places.

The paper is structured as follows. Section 2 is reviewing both theoretical contributions and empirical results from the literature. The region and the data are presented in Section 3, while the segregation pattern in the Oslo urban area is addressed in Section 4. We introduce the diversity measure in Section 5, discuss the modelling framework and some econometric issues. Section 6 presents the results following from alternative model formulations. Concluding remarks and important policy implications appear in Section 7.

2. Theory and empirical literature on how segregation and diversity affect local housing prices.

Following the capitalization hypothesis (Tiebout, 1956; Oates, 1969), property prices tend to be higher in areas with better amenities, better public services and perceived attractive social demographic neighbourhood compositions (Yinger, 1976; Brueckner et al. 1999). Segregation patterns by country background could be one important neighbourhood characteristic. On an aggregate city or regional level, the literature shows that increased immigration leads to higher housing prices; see, for instance, Saiz (2007) and Gonzales and Ortega (2013). However, there is less evidence on the impact that immigration may have on disaggregated local neighbourhood prices (Accetturo et al., 2014). Saiz and Wachter (2011) provide a pivotal contribution in this respect.

Accetturo et al. (2014) extend the work of Saiz and Wachter (2011). They develop a spatial equilibrium model, involving a utility function, which includes local amenities as one of the arguments. The level of local amenities is a function of the number of immigrants, and the differential effects of immigration on local housing values depend inter alia on the perception of these amenities and on native flight from areas with high immigrant density. Consequently, the sign of the relationship is an empirical question, and may vary across districts and cities. Arguments related to racial and religious issues, and to cultural diversity, for instance, may pull in different directions.

Immigration relates to both segregation and diversity. Issues related to local amenities and cultural capital are addressed in an emerging literature on whether diversity per se impacts well-being, attractivity and home prices. There are various answers to the question concerning which mechanisms may yield preferences for diversity. Diversity may harm the efficiency of communication and trust (Alesina and La Ferrara, 2005; Li, 2014) and thereby depress home prices. It may also provide a more diverse local supply of consumption possibilities (Bakens et al., 2013; Ottaviano and Peri, 2006). Hence, there are conflicting views on how diversity impacts the location decisions of natives, and consequently, the local housing prices. The net effect is thus an empirical question.

The economic literature on urban segregation is highly relevant for this paper. Cutler et al. (1999) focus on the "port of entry" mechanisms, which represent a voluntary self-selection whereby new immigrants choose to settle down close to earlier arrived members of the same group. Citing Borjas (1992), Wachter (2017, p. 495) also focuses on the tendency that "immigrants prefer immigrant-dense neighborhoods, due to the proximity they afford to people of the same national, ethnic, or linguistic group". Moreover, Amior and Manning (2018) note that it is well known that migrants are often guided in their location by the amenity of established co-patriot communities, for example, because of job networks or cultural amenities.

Inherent in much of the cited literature is the possibility that an inflow of immigrants to an area may induce a native flight to other districts in the urban area (Wachter, 2017), due to changes in housing prices, housing quality, and/or to how natives perceive the quality of local amenities. Hence, as Wachter (2017) points out, urban segregation tendencies do not solely reflect immigrant choices and the preferences of previous settlers. It may also correspond to what Cutler et al. (1999) denote "decentralized racism", and to natives who are willing to pay relatively more to live in primarily white areas.

According to Wessel and Nordvik (2018, p. 3) race could be perceived to be a proxy for a range of problems (e.g., poverty, crime and pollution) persisting in immigrant-dense areas. The so-called racial proxy hypothesis is also related to income and to the fact that those with higher incomes are able to buy a home in amenity-denser areas. Citing Ellen (2000), Wessel and Nordvik (2018) point to prejudiced race-based stereotypical perceptions of neighbourhoods among the native majority. These prejudices connect to neighbourhoods, not to individuals, and they may exist and be acted upon, even if, for instance, deprived neighbourhoods are significantly improved.

Our paper is primarily an extension of the scarce empirical literature in the field. Moye (2014) reviews this literature with a focus on whether the presence of minorities has a negative impact on housing prices in the USA. The overall conclusion, prior to 1965, is that race did not have a negative impact. However, Moye (2014) notes that there have been fewer such studies since 1988 and the results are mixed. In studies conducted after 2008, house values fell far more in neighbourhoods that were dominated by black people, whereas appreciation was higher than average in mixed black and white neighbourhoods. Using Australian data, Cobb-Clark and Sinning (2011) find that concentrations of immigrants with certain country backgrounds have a positive influence on local housing prices, while other groups have a negative or insignificant impact on housing prices. Based on data from Reading and Darlington (in the UK), Cheshire and Sheppard (1998) find a negative price elasticity with respect to the share of African-Caribbean population in a ward. In Italian cities, Accetturo et al. (2014) find that when local areas experience an inflow of immigrants and native flight, lower price growth follows.

Wachter (2017, p. 497) also highlights evidence supporting “a causal link between growing immigrant density, native flight, and a slowdown in housing value appreciation”. Using data from the USA and Italy, respectively, Saiz and Wachter (2011) and Accetturo et al. (2014) predict that the net effect of immigration is a deterioration of local amenities, and that housing prices will grow more slowly than average in areas hit by immigration. Finally, Wachter (2017) highlights the need to account for socio-economic variables, since the negative bivariate relationship between housing prices and immigrant density may reflect a tendency that immigrants are of lower socio-economic status. It is therefore important to account for socio-economic variables in the analysis to identify any links between housing prices and segregation.

Although our study is cross-sectional in nature, it should be noted that it also relates to studies of gentrification and housing price dynamics. This literature shows how neighbourhood socio-economic composition and housing prices—and the changes therein—interact. Guerrieri et al. (2013) and Delmelle (2017) are examples of this line of research. Based on examples from US cities, Hwang and Sampson (2014) highlight the danger that gentrification can increase racial segregation and inequality.

As noted, few empirical studies have explicitly explored the impact of immigrant segregation on housing prices, and even fewer have specifically investigated the impact of country background diversity on housing prices. One notable exception is Li (2014), who studied the impact of ethnic compositions in Vancouver, Canada, as measured by the Herfindahl index, from 1986 to 2001. One robust result in Li’s work is that neighbourhoods with relatively homogeneous minority populations are higher priced than neighbourhoods with more diverse minority compositions.

Bayer and McMillan (2008) argue that the diverse results from previous research are likely due to misspecifications, unobservable and unmeasurable features, and sorting. Accetturo et al. (2014) claim that the observed high degree of spatial autocorrelation among neighbouring districts represents a potentially serious drawback in the evaluation of their results. These issues thus motivate more empirical research on the capitalization of neighbourhood characteristics reflecting the country background of inhabitants.

3. The region and the data

According to Wessel (2016), Oslo was a city with pre-modern structures in the early 1980s, experiencing a massive suburban expansion. Since then, Oslo has been transformed into a rich, post-industrial city, with a substantial increase in the immigrant population. Thus, the study area has experienced some major changes. When Ottaviano and Peri (2006) needed an example of an ethnically homogeneous society, they referred to Norway in the 1960s. At that time, the situation in Oslo was similar to the rest of the country.

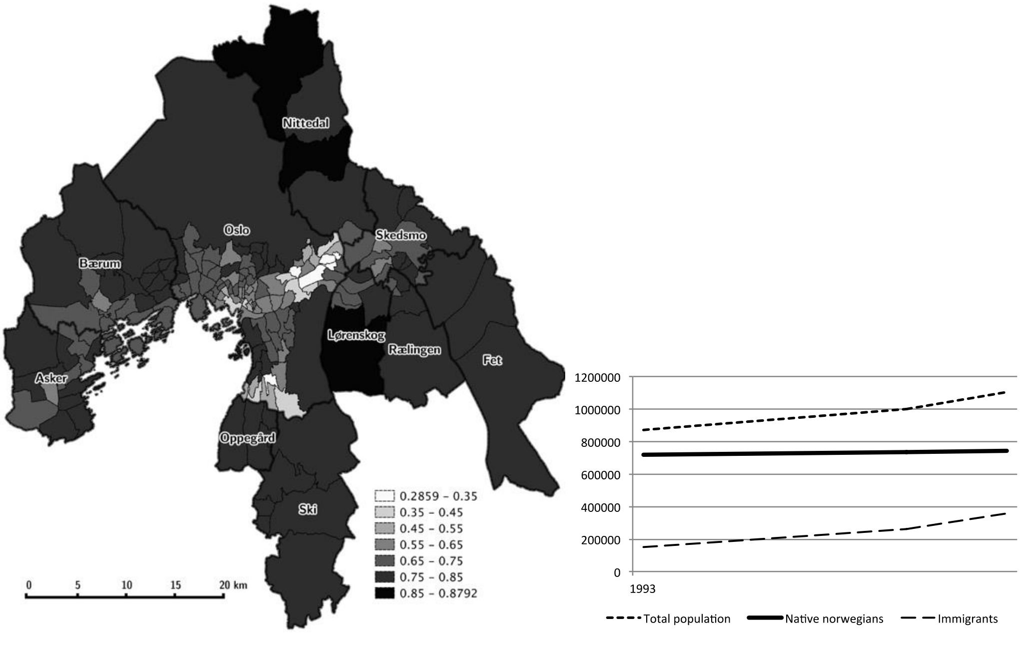
The Oslo urban area consists of 10 municipalities, with the municipality of Oslo in the centre. The total number of inhabitants in the area is over 1.1 million, which comprises about 20% of the Norwegian population. Figure 1a illustrates how the proportion of native inhabitants varied across the 182 subdistricts in 2011. We see a tendency for immigrants to be concentrated in the central parts of the Oslo urban area, with relatively lower proportions in the most peripheral and sparsely populated areas. Natives comprised 62.6% of the inhabitants of Oslo in 2011, while the corresponding figure was 75.8% for the aggregate of the other municipalities in the urban area. Figure 1a, indicates large variations between subdistricts within the Oslo municipality.

The graph in Figure 1b shows that the Oslo urban area has experienced a substantial population growth in the period 1993-2011, and that this growth roughly corresponds to the increased number of immigrants. Immigration is the main reason Oslo was one of the fastest growing metropolitan areas in Europe in this period. Note that we define immigrant status according to an individual and his/her parents' country of birth. An individual born in Norway to, for example, Indian parents is counted as Indian, while anyone born in Norway to at least one Norwegian parent is defined as native.

Our analyses utilize pooled time series and cross-sectional data from different sources, for the period 2009-2012. There are 99,852 observations of individual completed housing transactions, 73% of them from the municipality of Oslo. The price data are taken from the net-based housing sale portal, Finn.no, which covers around 70% of the sales in the area.

A neighbourhood is defined by the clusters of census tracts making up the 182 subdistricts. The average number of people in these districts was 6531 in 2011, and the maximum was 16484. The composition of the population in the neighbourhoods is measured along many dimensions. Public population registers provide information on mobility, age and country background of individuals, while the register of education gives educational achievements. Tax registers are the source of information on income distribution. Finally, information on receipt of welfare benefits is taken from the registers of the Central Welfare Agency of Norway. The socio-economic and demographic data are from 2011, and the lagged data to be used in the empirical analysis are from 2006. Hence, we use a wide array of socio-economic characteristics, based on high-quality register data with complete coverage, which further allows us to use a very fine-grained definition of neighbourhoods.

All included distances are measured by the shortest travelling time by car, between postal delivery zones. We account for speed limits, but not for the within-zone travelling time. In calculating values of the labour market accessibility, we use employment data from Statistics Norway .We regard data from all these sources to be of very good quality. For more details on the data, and on the region, see Nordvik and Osland (2016).



a) The Oslo urban area b) Population growth, 1993-2011

Figure 1: a) The proportion of native inhabitants in the 182 subdistricts in 2011. b) Population growth of native Norwegians and immigrants, based on observations from 1993, 2006 and 2011.

4. The segregation pattern in the Oslo urban area

Table 1 shows that the five largest groups of immigrants in 2011 come from Sweden, Pakistan, Poland, Denmark and Somalia. Thus, the population in the study area has become very diverse in terms of country background. Some of the subdistricts have people from more than 130 countries. The last two columns in Table 1 provides information on how inhabitants of different country background are spread throughout the Oslo urban area, measuring the percentage of a zone where a specific group is significantly represented: the meaning of “significantly represented” will be explained and discussed in Section 5.1, where we address definitions of diversity.

First, we provide some descriptive material on observed segregation tendencies in Oslo. One common measure of the tendency that a group is segregated in a geography is given by the index of dissimilarity, (*ID*), which is defined by:

$$ID = 0{,}5 \sum_m \left| \frac{P_{im}}{P_i} - \frac{P_{jm}}{P_h} \right|,$$

where P_{im} is the number of immigrants from group *i* residing in zone *m*, P_i is the number of immigrants from country *i* in the whole area, P_{jm} is the number of people from other countries than *i* residing in subdistrict *m*, while P_h is the total population in the area with a background from other countries than *i*. If the proportion of immigrants from different countries is the same in all the zones of the geography, then *ID* = 0. The index is calculated from information in 182 subdistricts, and the value will be higher the more segregated the residential location pattern is for a specific group of immigrants.

Table 1 reveals a tendency that immigrants from neighbouring Western countries are more integrated than immigrants from developing countries. Notice that the large group of immigrants from Pakistan has been largely and consistently segregated, while for instance, Iranians have become more integrated over time. The map in Figure 2a visualizes of the residential location pattern of immigrants from Pakistan within the municipality of Oslo in 2011. We have omitted from the maps in Figure 2 some subdistricts that covers relatively large wilderness areas within the municipality of Oslo, but with very few inhabitants.

Table 1. Number of inhabitants, values of the dissimilarity index (ID), and percentage number of subdistricts where inhabitants with different country background are represented.

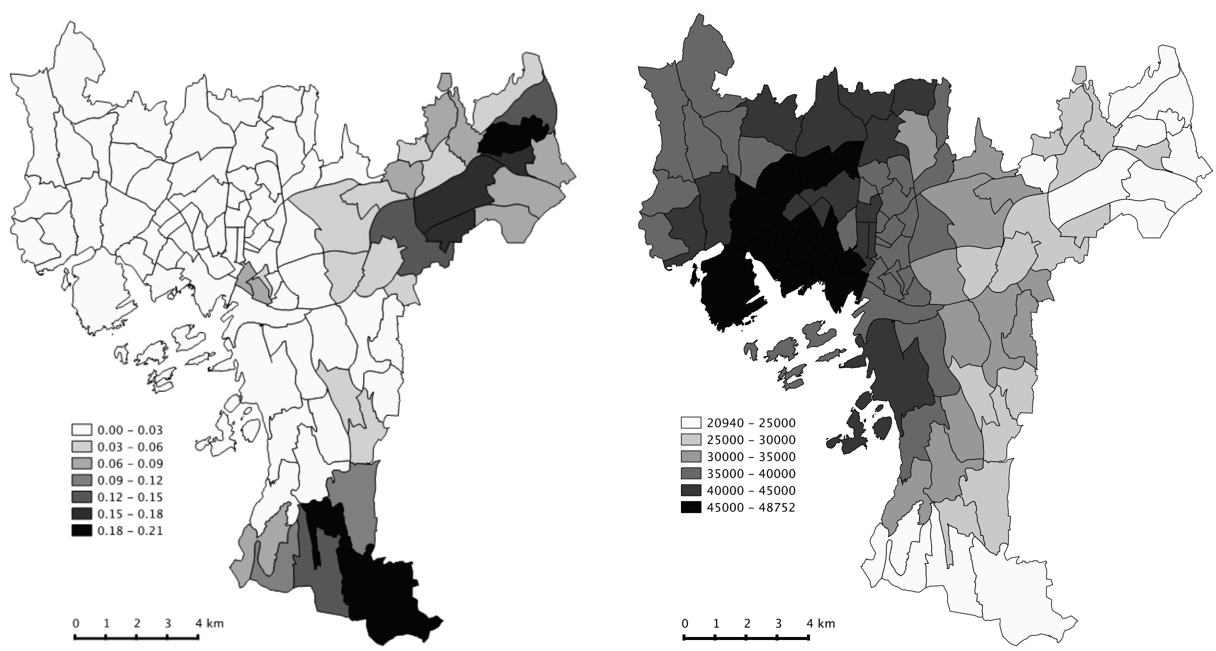
Country background	*#1993*	#2006	#2011	*ID* 1993	*ID* 2006	*ID* 2011	SUB, k=0.1	SUB, k=1
Norway	718680	731800	745386	0.1910	0.2170	0.2214	99%	70%
Denmark	15387	17012	17353	0.1080	0.1089	0.1075	91%	45%
Sweden	15037	25404	34781	0.1563	0.1658	0.1879	92%	39%
Germany	6992	9239	11865	0.1920	0.1825	0,1666	83%	42%
Poland	3099	6202	20493	0.2206	0.1601	0.1655	83%	38%
Great Britain	9529	11148	13035	0.2460	0.1898	0.1856	86%	45%
USA	9626	10439	11540	0.2444	0.2473	0.2434	83%	42%
Somalia	1304	10333	15888	0.5769	0.4608	0.4855	37%	17%

Sri Lanka	2668	8895	10128	0.5425	0.5813	0.5653	40%	18%
Iran	2924	8356	10245	0.4097	0.2833	0.2464	64%	32%
Pakistan	17048	26754	31174	0.5149	0.5054	0.5130	50%	38%
Thailand	890	2742	4041	0.3298	0.2487	0.2186	56%	33%
Turkey	4422	7687	8911	0.4947	0.4588	0.4486	47%	19%

The zonal level for ID is subdistricts. The last two columns use data from 2011, and we use the zonal level applied in the empirical analysis reported in Table 2, which is a census tract. For a definition of *k*, see Section 5.1.

Figure 2a reflects a marked tendency of inhabitants with country background from Pakistan being concentrated in the eastern parts of Oslo. This is a representative picture of the residential location pattern for many groups of immigrants from developing countries, and coincides with Oslo being a divided city. Traditionally, the wealthy middle class have lived in the western parts of Oslo, with the working class and less affluent groups of the population traditionally residing in the eastern parts.

There has been a sharp increase in observed housing prices in the Oslo urban area, which, according to Wessel (2016, p. 137) is “hardly met elsewhere in Western Europe”. In line with the observations reported by Gonzalez and Ortega (2013), the increase in housing prices has not been the same across different subdistricts and neighbourhoods in the geography. Figure 2b illustrates how housing prices vary over space in the municipality of Oslo, and supports to the claim that Oslo is to some degree a divided city, with a marked tendency that western parts have the highest housing prices. Comparing Figure 2a and 2b, there appears to be a close inverse relationship between the proportion of immigrants from Pakistan and the average housing price per m^2 in the subdistricts, and a similar relationship exists for other groups of immigrants from developing countries. Indeed, the correlation coefficient of -0.7223 between housing prices and the proportion of inhabitants in the municipality of Oslo with country background from Africa or Asia is highly negative. That said, despite some distinct patterns of segregation in terms of socio-economic characteristics, Musterd (2005) finds that Oslo has a relatively low level of ethnic segregation compared with many other European cities.



a) Country background from Pakistan, 2011 b) Housing prices, 2009/10

Figure 2. a) Proportion of inhabitants with country background from Pakistan in 2011. b) housing prices per m^2 in 2009/2011 in subdistricts of the Oslo municipality.

5. The modelling framework

In this section, we first introduce the diversity index that is used. Second, we specify the linear predictor and the link function, connecting the linear predictor to the dependent variable. Then, we introduce spatial random effects, in terms of a conditional autoregressive mechanism, before we address possible endogeneity problems.

5.1 Measuring the diversity of country background

Our paper explores and tests empirically whether the presence and residential location patterns of inhabitants with different country background have an impact on the attractivity of neighbourhoods, and are thereby capitalized into local home prices. It is far from obvious how diversity should be measured, and there exist various approaches in the literature. A commonly used way to measure spatial variations in diversity of different characteristics is the regional diversity index (RDI) (see e.g., McCann (2013)). In measuring the local diversity of country backgrounds, the regional diversity in subdistrict *j* is defined by:

$$RDI_j = \frac{1}{\sum_{i=1}^{219}\left|\frac{A_{ij}}{A_j} - \frac{A_{iR}}{A_R}\right|},$$

where A_{ij} is the number of inhabitants with country background *i* in subdistrict *j*, A_j is the total number of inhabitants in subdistrict *j*, A_{iR} is the number of inhabitants with country background *i* in the region, and A_R is the total number of inhabitants in the region. In the Oslo area, there are observations of 219 different country backgrounds.

We introduce an alternative measure of diversity, that is basically defined by counting the relative number of country backgrounds that are significantly present in a neighbourhood. The following notation is introduced for a more precise definition of the measure.

NC_{ir} = the number of inhabitants with country background i in zone r

N_r = the number of inhabitants living in zone r

NC_i = the number of inhabitants with country background i in the region

N= the number of inhabitants in the region

We further define $nc_{ir} = \frac{NC_{ir}}{N_r}$ and $nc_i = \frac{NC_i}{N}$, and introduce the location quotient, $LQ_{ir} = \frac{nc_{ir}}{nc_i}$. Let k represent a threshold value of LQ_{ir}, used to define whether a specific group of immigrants is present in a zone. The group of immigrants is defined to be significantly present in zone r if $LQ_{ir} > k$. For k = 1.0, the relative presence of a group of immigrants in the local population must be no lower than the average representation of this group in the regional population. If k = 0.5, a group of immigrants is defined to be significantly represented in the zone if the local proportion is at least half the size of the regional proportion. Hence, the presence of a group can be represented by the following indicator function:

$$IND(ir) = 1 \quad if \quad LQ_{ir} \geq k$$

A corresponding index of diversity can then be defined by

$$DIV_k(r) = \frac{1}{m}\sum_{i=1}^{m} IND(ir)$$

This index takes a value of 1 if all the immigrant groups are significantly represented in the zone; it takes a value of 0.5 if 50% of the groups are significantly represented.

We introduce the parameter k as a device to control how strictly one should set the requirement for a group to be significantly present in a sub-population. The last two columns of Table 1 provide information on the percentage number of subdistricts where inhabitants of a specific country background are represented. This information is given for a selection of country backgrounds and for two values of k. It follows from Table1 that the number of subdistricts where a group is present is substantially higher for the low value of k. This is according to intuition, and means that a low value of k produces a high value of the diversity measure, *DIV*. For k = 1.0, the subdistrict with a median percentage has a value of *DIV* equal to 9.1, while the corresponding median value is 13.2 in the case of k = 0.1.

There is no common sense regarding what is a reasonable value of k, and the choice of k may seem to be subject to a certain arbitrariness. However, in our revealed preference approach, k can be estimated to reflect what kind of neighbourhood that is preferred by the agents in the housing market. In principle, we can estimate *DIV* to contribute negatively to housing prices for a high value of k, but positively for a low value of k. If so, this reflects a situation where house buyers are attracted by neighbourhoods with a thin representation of people from a wide number of countries. At the same time, however, such a result would pull in the direction of low local housing prices in a neighbourhood with relatively large groups of immigrants from this wide range of countries.

From a theoretical point of view, it is difficult to recommend a specific measure of diversity. Our pragmatic approach is to leave it to empirical investigation. In our case, the measure *DIV* consistently results in a better model performance than the *RDI* measure. This applies for all the model formulations that we have considered, and for reasonable values of k. Hence, we have chosen not to report the results based on model formulations using *RDI*. However, we do not claim that *DIV* is in general superior to *RDI*. In the problem we are studying, there are many possible country backgrounds involved. Hence, *DIV* can be assigned a high number of values, allowing us to distinguish between zones on a relatively continuously defined opportunity set. However, when using data with only a few categories, *RDI* may indeed be a more appropriate measure of diversity.

For our purpose, an appealing feature is that *DIV* allows a degree of flexibility in estimating k, and in deciding whether a thin representation of immigrants is preferred to a thicker representation in a neighbourhood. Like other diversity measures, *DIV* incorporates information on multiple groups simultaneously, making direct interpretation challenging. Nevertheless, we believe that DIV is based on an intuitively appealing counting routine.

5.2 Basic general linear model

As a first element of a general linear model (GLM), the stochastic dependent variable (house price) is assumed to follow an exponential family of probability distribution. In our data, the distribution of house prices was found to be closely approximated by the normal distribution, which means that this part of the model is equivalent to the ordinary multiple regression model. The second element is the linear predictor, which will be discussed in more detail below; the third element is represented by a log link:

$$E(\log P_i) = \log \mu_i = \eta_i \qquad (1),$$

where P_i is the price of house i in a particular year, and η_i is the linear predictor.

The specification of the linear predictor introduces the set of variables that are hypothesized to influence house prices. The individual variables can be represented by the following vectors:

DWELLING	attributes of the observed dwelling
YEARDUMMIES	year dummies for 2010, 2011, and 2012
LOCATION	spatial characteristics related to the location of an observation
SOCIOECDEM	economic and sociodemographic characteristics of the population
COUNTRY	the proportion of the population with the respective country background
DIVERSITY	the diversity of the population, with respect to country background

The reference categories for the dummy variables are: apartments, construction year before 1940, a freeholder's house, and houses sold in 2009. The linear predictor for observation i is formulated as:

$$\eta_i = \beta_0 + \beta_1(\log \mathbf{DWELLING})_i + \beta_2(\log \mathbf{LOCATION})_i + \beta_3(\log \mathbf{SOCIOECDEM})_i + $$

$$+\beta_4(\log \mathbf{COUNTRY})_i + \beta_5(\log \mathbf{DIVERSITY})_i + \sum_{t=10}^{12} \beta_1(\log \mathbf{YEARDUM})_i + \in_{it} \quad (2)$$

Some of the variables are related to the notion of social capital, which is measured, for instance, as the length of stay in the neighbourhood, the percentage of single parent households, the percentage of married couples with children and the degree of home ownership. A list of the individual variables for each category is provided in the Appendix.

5.3 A conditional autoregressive (CAR) spatial model formulation

Rather than just introducing a set of zone-specific dummy variables in a spatial fixed effect model formulation, the CAR spatial model accounts for the possibility that the intercept changes smoothly, or continuously from zone to zone. The zone-specific random effects, γ_i, means that the linear predictor is extended:

$$\eta_i = \beta_0 + \sum_{j=1}^{k} x_i\beta_j + \gamma_i \qquad (3),$$

where x_i are elements in the matrix of covariates, defined in Equation (2). The 182 random effects are defined at the subdistrict level, and a subdistrict is a cluster of census tracts. The connectivity and neighbour structure of the zones are introduced by the weight matrix **W**, with elements defined by $w_{ij} = \frac{a_{ij}}{a_i}$, where

$$a_{ij} = \begin{cases} 1, & \text{if zone } i \text{ and zone } j \text{ are neighbours} \\ 0, & \text{otherwise} \end{cases}$$

and $a_i = \sum_j a_{ij}$ denotes the number of zones that are neighbours to zone i. This weight matrix is row-standardized, $\sum_j w_{ij} = 1$, with weights defined according to queen-based contiguity (see, e.g., Anselin (1988)). The random effects are assumed to be distributed according to the multivariate normal distribution, defining a Gaussian CAR model. The model allows for local spatial smoothing. Let $\gamma_{-r} = (\gamma_1, \dots, \gamma_{r-1}, \gamma_{r+1}, \dots, \gamma_n)$, that is, the vector of all elements in γ, except the *r*th element. The conditional distribution of the random effects $(\gamma_r|\gamma_{-r})$ is then given by:

$$\gamma_r|\gamma_{-r} \sim N\left(\frac{\sum_{s=1}^{n} w_{rs}\gamma_s}{\sum_{s=1}^{n} a_{rs}}, \frac{\tau^2}{\sum_{s=1}^{n} a_{rs}}\right) \qquad (4)$$

The scalar τ reflects the general variance of the random effects. As pointed out in Osland et al. (2016), the model is "improper", because the covariance matrix of the simultaneous distribution is not positive definite. However, it is still being used (Gschlößl and Czado, 2007). Bivand et al. (2017) demonstrate that this version of a spatial autoregressive model gives a reasonable and satisfying explanation to housing market data, compared with alternative autoregressive approaches.

The CAR model used in this paper is based on a Bayesian hierarchical model formulation of the random effects. It follows from Equations (3) and (4) that the conditional expectation of house prices depends on the average of the random effects in the neighbouring zones, while the conditional variance is inversely proportional to the number of neighbouring zones.

5.4 Endogeneity issues

Our data shows a strong negative correlation between housing prices and the proportion of inhabitants with country background from Africa and Asia in the city of Oslo (see Section 4). This negative correlation may be given a causal interpretation (Saiz and Whachter, 2011; Accetturo et al. 2014)), but not necessarily so. Obviously, the reasons for this uncertainty are a range of potential endogeneity problems, which we address in the following paragraphs.

First, the mentioned negative correlation could relate to confounding factors. Saiz and

Wachter (2011) conclude that this feature could occur for other reasons than "foreignness per se" (p. 187). A local neighbourhood concentration of inhabitants with unfortunate socio-economic characteristics may also have a negative impact on housing prices. For example, in our data, we find a correlation coefficient of -0.65 between the proportion of people from Africa and Asia and median work-related income in neighbourhoods. The correlation coefficient is 0.71 between the proportion of this group and the local proportion of people with low education. Ignoring such correlations introduces a potential negative bias in the parameter estimates representing the impact of foreign country background. Likewise, excluding socio-economic variables may cause a positive bias in the estimated coefficient representing the impact of natives. To separate and unbundle the impact of country background on housing prices as far as possible, we have incorporated demographic and socio-economic characteristics of census tracts into the model.

Reverse causality could be another potential source of endogeneity. Groups of immigrants may choose certain neighbourhoods because housing prices are lower than in other areas. Saiz and Wachter (2011) successfully address the issue of reverse causality between changes in average housing values in neighbourhoods and the single variable immigrant density. As an instrument, they use a spatially lagged gravity-based measure of proximity to other immigrants. This variable is found to be exogenous. Based on the results from Hausman tests, they conclude that the ordinary least squares estimator is to be preferred. However, it can be argued that the need for a spatial type of instrument like the one used by Saiz and Wachter (2011) is less urgent in our case. Such a contagion effect is to some degree captured by the random effects in the CAR formulation. In general, the CAR model accounts for omitted systematic spatial variation in relevant characteristics, and contributes to removing problems related to reverse causality involving a spatial dimension.

We have also experimented with an instrumental variable approach. However, in contrast to, for example, Saiz and Wachter (2011) we have a multiple of potentially endogenous variables. We focus on proportions of people with different types of country background, not just immigration per se. As explained in Baum et al. (2003) this creates a range of problems including difficulties in testing for exogeneity and the IV estimator could be biased and inconsistent. We have applied the 2SLS estimator using variables lagged in time as instruments. The lagging implies using 1993 or 2006 values of the relevant variables, respectively. However, the results were very unstable, mainly because of the weak instruments related to the potential of multiple endogeneous variables. We have therefore chosen not to report the results based on this estimator.

In general, time-lagged variables as a proxy for the endogenous variable is frequently used to deal with simultaneity due to reverse causality (Reed, 2015). This may be an appealing idea since the dependent variable certainly does not cause the lagged value of the endogenous variable. In addition to the spatially lagged density measure of immigration, Saiz and Wachter (2011, p. 175) also claim that, for example, time-lagged socio-economic variables in levels, as a place-specific characteristic, could be predictors of forthcoming housing prices.

Given the above discussions and the difficulties in dealing with this type of endogeneity, we have chosen to experiment extensively with time-related lagging of variables representing socio-economic, demographic, and country background characteristics. This approach does not enable us to test whether we have endogeneity problems. Moreover, Reed (2015) emphasizes that our approach will not necessarily solve the simultaneity problem, for instance, because the time-lagged variables frequently are strongly autocorrelated. However, obtaining parameter estimates with signs according to a priori expectations, and considering the goodness of fit and the stability of the results across different model formulations, provide some indications of the existence and significance of an endogeneity problem.

5.5 Estimation method and criteria for model selection

Since the likelihood function, resulting from the CAR model is high-dimensional and complex, Bayesian inference and MCMC simulations are used as an estimation approach. Estimation is carried out by a combination of the Metropolis-Hastings and Gibbs algorithm (see, e.g., Korn et al. (2010)), using the WinBUGS software (Spiegelhalter et al., 2003), with a burn-in of 5000. For model selection, we use the Bayesian analogue to the Akaike Information Criterion (AIC), namely the Deviance Information Criterion (DIC), (see e.g., Gelman et al. (2014)). DIC approximates AIC, and models with a low value of DIC are preferred. Moran's I of the random effects is also calculated and reported to provide information on the presence of spatial dependencies and smoothing.

6. Results

This section presents and discusses the estimation results. Section 6.1 focuses on model selection, while section 6.2 briefly addresses the control variables. We primarily focus on how diversity and concentrations of different groups of immigrants affect local housing prices, and these questions are discussed in Sections 6.3 and 6.4.

6.1 Model selection

Theoretical arguments do not contribute sufficiently to constraining the set of possible configurations of country background in the model specification. Should we utilize a fine-grained specification with dummies for each individual country background or should we aggregate into groups of countries? We have relied on testing a number of possible ways to aggregate, for example, by development status and continents. Using DIC values (see Section 5.5) is an appropriate way to rank the performance of non-nested models with differing degrees of freedom. Based on this, we found that an aggregation into nine country groups (some of them containing one single country) account appropriately for country background. In the remainder of the paper, we will refer to this as a more disaggregated set (DA). In addition, we present a more parsimonious model that aggregates country background into two groups (A).

Next, we address the question of whether to use contemporaneous (2011) or time-lagged (2006) values of the shares from each (aggregate of) country background and other socio-economic characteristics of neighbourhoods. Comparisons of DIC values reveal a better fit when lagging compared with no lagging. However, an even better model is the one that uses contemporary values of the socio-economic controls, and lags the variables related to country background and segregation. We continue our presentation of results using the model with the highest DIC value. The coefficients we are primarily interested in (i.e. shares from different country backgrounds and our diversity measure), are quite similar in these last two specifications. Table 2 presents estimation results from five different specifications:

M1(A) an aggregated set of covariates representing country background, no lagged covariates

M2(A) an aggregated set of covariates representing country background, lagged values of the covariates representing country background

M1(DA) a more disaggregated set of covariates representing country background, no lagged values of covariates

M2(DA) a more disaggregated set of variables representing country background, lagged values of the variables representing country background

M3(DA) a more disaggregated set of variables representing country background, lagged values of the variables representing country background, and a definition of country background diversity corresponding to a value of $k = 1$

The results presented in Table 2 are based on a diversity measure calculated for $k = 0.1$, representing a rather thin definition of diversity. For comparison, we have included complete estimation results for a model M3(DA) using a thick definition of diversity ($k = 1$). Later, we demonstrate how the estimated effect of diversity and the overall model performance vary as k is varied. Table 3 provides definition of variables.

6.2 Control variables

The set of control variables includes descriptions of the individual housing units (vintage, size and house type), location (distance from the City Hall in different directions), a measure of work place accessibility (Osland and Thorsen, 2008), and a set of variables describing the socio-economic composition of the neighbourhood. Among the neighbourhood composition controls, we include moments of the local income distribution (calculated for male prime-age residents), housing turnover rates, take-up rates of different welfare benefits, education and age. We also include compositions of the local housing stock according to house types.

The signs and magnitude of the coefficients of the controls agree with prior expectations and earlier studies. One particular result of notice for Oslo is that travelling time to the CBD in the direction north depresses home prices significantly, while the effect of travelling time to the west does not have a significant effect at the 5 % level of significance (see Table 4).

The income of the median male income earner in a neighbourhood actually has a substantial and significantly positive impact on housing prices. This accords with, for example, Cobb-Clark and Sinning (2011), who explain the results by claiming that there are positive externalities when residing in neighbourhoods with many wealthy people. Note that the standard deviation of incomes for male income earners also has a strongly positive impact on housing prices. This may reflect a tendency that the market appreciates heterogeneity with respect to income, rather than homogeneity.

6.3 The impact of concentrations in country background

The general finding in Table 2 is that the shares of inhabitants from specific (aggregates of) countries have a statistical significant impact on home prices. However, in part this impact is not of any substantial magnitude. Inhabitants with country background from Turkey and Central Asia have a statistically significant negative impact on local housing prices. The negative influence is slightly stronger, but is still small, for developing countries outside Central Asia. Concentrations of inhabitants from Western countries are estimated to have a significantly positive impact on local housing prices.

As a main result, concentrations of different groups of immigrants are estimated to be only very weakly associated with partial local variations in housing prices. Partially increasing the share of inhabitants with country background from Turkish or Central Asian with 10 percentage points depresses predicted housing prices by less than 0.5%.

The most distinctive result is that variations in local housing prices are positively related to the proportion of native residents in a neighbourhood. This very significant effect implicitly means that there is a relatively strong negative impact related to the presence of non-natives in a neighbourhood.

When comparing the results based on the models M1(DA) and M2(DA), it follows that lagging the country background variables in time only has a marginal impact on individual parameter estimates. This is an indication that most of the results can be interpreted in causal terms. Based on significantly strong positive estimates, we are confident that, for example, a high proportion of native residents indeed contributes to higher housing prices. From a theoretical point of view, there is no reason to believe that high local housing prices by themselves contribute to attracting native residents. Hence, reverse causality and simultaneity bias are not expected to represent a problem in this case.

Figure 3a illustrates how variations in the proportion of native Norwegians in a neighbourhood affect the local house prices. The vertical axis measures the predicted price for a standard house, which is defined as a semi-detached house built in the 1970s, sold in 2012, and with average values of dwelling attributes. We have used parameter estimates based on model M2(DA), ignoring the random effects. The curve in Figure 3a is defined for all values in-between the minimum and maximum observed values of the proportion of native Norwegians in a census tract. The estimated parameter value of 0.07029 means that the price of this standard house is predicted to be about NOK 500,000 (about 18%) more expensive in the neighbourhood with the maximum observed proportion of native Norwegians than in the neighbourhood with the lowest observed concentration. Notice that the effect of a marginal increase in the proportion of native Norwegians is predicted to be a decreasing function of this proportion.

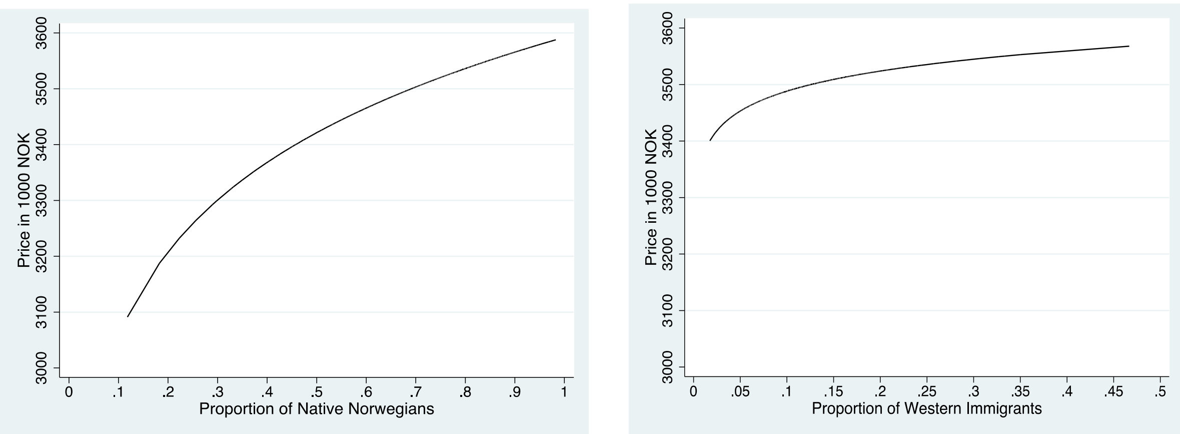
The horizontal axis in Figure 3b refers to variations in the proportion of Western immigrants. Observed variations in this proportion contribute to explaining variations in house prices of about NOK 150,000. It follows from the results referring to M2(DA) in Table 2 that this prediction is based on a parameter estimate of 0.01464.

Table 2: Results based on alternative model formulations.

	M1(A)	M2(A)	M1(DA)	M2(DA)	M3(DA)
Constant	9.3^{***}	9.129^{***}	9.078^{***}	9.409^{***}	9.335^{***}
	(0.1711)	(0.1685)	(0.1728)	(0.1706)	(0.1704)
AGE1019	-0.005142^{**}	-0.00503^{**}	-0.006457^{**}	-0.001338	-9.443E-4
	(0.002539)	(0.002529)	(0.002545)	(0.002543)	(0.002544)
AGE2030	-0.01467^{***}	-0.01613^{***}	-0.01371^{***}	-0.01509^{***}	-0.01009^{***}
	(0.003591)	(0.00356)	(0.003614)	(0.003577)	(0.003617)
AGE3040	0.04638^{***}	0.04228^{***}	0.04362^{***}	0.04444^{***}	0.04534^{***}
	(0.006015)	((0.005943)	(0.006063)	(0.006034)	(0.006044)
AGE4050	-0.03188^{***}	-0.03343^{***}	-0.03047^{***}	-0.03046^{***}	-0.03443^{***}
	(0.005882)	(0.005836)	(0.005879)	(0.005862)	(0.005863)
AGE5060	0.02951^{***}	0.03021^{***}	0.02647^{***}	0.02854^{***}	0.0283^{***}
	(0.004551)	(0.004547)	(0.004525)	(0.004597)	(0.004599)
AGE6070	0.01509^{***}	0.0124^{***}	0.01351^{***}	0.01223^{***}	0.01267^{***}
	(0.003094)	(0.003086)	(0.00313)	(0.003089)	(0.00309)
AGE70+	0.002767^{*}	0.006098^{***}	0.001751	0.008542^{***}	0.009243^{***}
	(0.001518)	(0.001533)	(0.001525)	(0.001541)	(0.001539)
RDI_{age}	-0.02169^{***}	-0.02472^{***}	-0.02294^{***}	-0.02458^{***}	-0.02426^{***}
	(0.002784)	(0.00278)	(0.002801)	(0.002821)	(0.002823)
INC(0711)MED	0.09212^{***}	0.1043^{***}	0.1004^{***}	0.09895^{***}	0.1025^{***}
	(0.01198)	(0.01183)	(0.01232)	(0.01194)	(0.01193)
INC(0711)SD	0.07704^{***}	0.07826^{***}	0.07827^{***}	0.07293^{***}	0.07393^{***}
	(0.0024)	(0.002395)	(0.002403)	(0.002417)	(0.002414)
HIGHEDU	-0.002505	0.001663	-0.002815	-0.001013	0.00172
	(0.005711)	(0.005673)	(0.005684)	(0.00574)	(0.00574)
LOWEDU	-0.004673^{**}	-0.006427^{***}	-0.006689^{***}	-0.007869^{***}	-0.00605^{***}
	(0.00206)	(0.001999)	(0.00199)	(0.001963)	(0.001971)
SOCSECBEN	0.001143^{**}	6.763E-4	7.672E-4	6.456E-4	0.001324^{**}
	(5.504E-4)	(5.49E-4)	(5.657E-4)	(5.466E-4)	(5.413E-4)
5OUTMOVE	0.001596^{***}	0.001972^{***}	0.001903^{***}	0.00358^{***}	0.004377^{***}
	(4.841E-4)	(4.82E-4)	(4.822E-4)	(7.11E-4)	(7.074E-4)
OUTMOVE	0.02202^{***}	0.01895^{***}	0.02039^{***}	0.01459^{***}	0.0119^{***}
	(0.003966)	(0.003949)	(0.003988)	(0.004008)	(0.004007)
INCGT120	-0.1519^{***}	-0.1381^{***}	-0.1373^{***}	-0.1443^{***}	-0.146^{***}
	(0.01256)	(0.01226)	(0.0126)	(0.01241)	(0.01243)
CASHCARE	-0.002182^{***}	-0.002453^{***}	-0.002413^{***}	-0.001981^{***}	-0.001356^{**}
	(6.571E-4)	(6.558E-4)	(6.65E-4)	(6.609E-4)	(6.585E-4)
NATSECBEN	-0.02498^{***}	-0.02305^{***}	-0.02392^{***}	-0.02257^{***}	-0.02279^{***}
	(0.002268)	(0.002262)	(0.002267)	(0.002247)	(0.002248)
DETPROP	-0.001192^{***}	-0.001811^{***}	$-8.896E-4^{**}$	-0.001161^{***}	$-9.763E-4^{**}$
	(3.954E-4)	(4.029E-4)	(3.972E-4)	(4.041E-4)	(4.038E-4)
APAPROP	0.01318^{***}	0.01281^{***}	0.01377^{***}	0.01058^{***}	0.009949^{***}
	(0.001789)	(0.001783)	(0.001828)	(0.001799)	(0.001797)
BLOCKPROP	-4.835E-4	$-8.465E-4^{**}$	-2.55E-4	$-6.908E-4^{*}$	-4.87E-4
	(3.992E-4)	(4.0E-4)	(4.014E-4)	(3.988E-4)	(3.987E-4)
SMALLPROP	-0.001041	5.472E-4	-0.001538	0.001208	0.001873
	(0.001312)	(0.001316)	(0.001325)	(0.001318)	(0.001318)
NORWEGIAN	0.06431^{***}	0.06422^{***}	0.07488^{***}	0.07029^{***}	0.06866^{***}
	(0.008334)	(0.007957)	(0.008045)	(0.007801)	(0.007814)
POLAND			-0.001639^{***}	2.704E-5	3.58E-4
	()	()	(4.244E-4)	(2.91E-4)	(2.894E-4)
RUSSIA			$7.431E-4^{***}$	6.293E-5	$4.732E-4^{**}$
	()	()	(2.105E-4)	(2.225E-4)	(2.212E-4)
OCEANIA			$9.762E-4^{***}$	0.001044^{***}	0.001285^{***}
	()	()	(1.908E-4)	(1.666E-4)	(1.659E-4)
SOMALIA			0.001025^{***}	$-5.128E-4^{**}$	-2.319E-4
	()	()	(2.403E-4)	(2.225E-4)	(2.214E-4)
TURKEY			-0.00111^{***}	-0.00185^{***}	-0.001551^{***}
	()	()	(2.06E-4)	(2.294E-4)	(2.295E-4)
CENTRALASIA			-0.00193^{***}	-0.001641^{***}	-0.001424^{***}
	()	()	(3.301E-4)	(2.804E-4)	(2.792E-4)
DEVELOPING			-0.001958	-0.005757^{***}	-0.005142^{***}
	()	()	(0.001221)	(6.736E-4)	(6.766E-4)
WESTERN			-0.01581^{***}	0.01464^{***}	0.01827^{***}
	()	()	(0.002743)	(0.003391)	(0.003425)
NON-WESTERN	-0.01665^{***}	-0.01035^{***}			
	(0.002433)	(0.001281)	()	()	()
$DIV_{0.1}$	0.02349^{***}	0.02169^{***}	0.02709^{***}	0.02732^{***}	
	(0.003161)	(0.00313)	(0.003808)	(0.003577)	()
$DIV_{1.0}$	-	-	-	-	-2.269E-4
	(-)	(-)	(-)	(-)	(0.003634)
τ	39.88^{***}	39.49^{***}	39.47^{***}	39.46^{***}	39.48^{***}
	(4.664)	(4.599)	(4.556)	(4.557)	(4.556)
DIC	-43820.700	-43857.800	-43917.000	-44070.700	-44006.400
Morans' I	0.768244	0.767430	0,771818	0,770945	0,774413

Note: * significant at the 0.1 level, ** significant at the 0.05 level, *** significant at the 0.01 level; robust standard errors in parentheses.

The discussion reveals that there are relatively few examples where simultaneity and reverse causality are expected to represent serious problems in interpreting the estimation results. A large range of control variables is incorporated, and the CAR formulation accounts for omitted spatially correlated information on characteristics at the zonal level. This also contributes to justifying an interpretation of the estimates that represent the partial impact of the covariates on housing prices, rather than interpreting the parameter estimates in terms of mere correlations, capturing the effects of omitted neighbourhood information. As noted, the strong effect related to the presence of natives in a neighbourhood, implicitly means that there is a corresponding negative effect related to the presence of non-natives. However, another main conclusion is that there are no strong tendencies indicating that immigrants from specific countries are not wanted as neighbours by house buyers.



a) Native Norwegians

b) Western immigrants

Figure 3: The effect of observed neighbourhood variation in the proportion of a) native Norwegians and b) Western immigrants on the price of a standard house.

Conclusions are even more nuanced when one considers neighbourhoods as closed sets of neighbours. Sums of shares are obviously bounded from above at unity. The partial effects we have discussed tell us much about the structures and variations in house prices, but they are still partial. The effect on prices of an increase in the share of inhabitants from, for example, Central Asia by 10 percentage points depends on whom the new inhabitants replaces. If they replace any of the other groups with a non-native country background (including Western), the effect is virtually zero. If a similar increase in the share of Central Asians is accompanied by a corresponding decrease in the share of natives, house prices drop by 1.4%. Hence, any claim that the Oslo housing market displays no major immigrant neighbour penalty in house prices relies on the partiality of the arguments. See also Accetturo et al. (2014) for a discussion of similar issues.

6.4 Impact of diversity in country background

With a significant positive coefficient of 0.0273, diversity in terms of country background feeds positively and significantly into house prices. This gives a highly non-linear association between diversity and prices of housing. Figure 4 uses the same definition of a standard house as in Figure 3. It illustrates how variations from the minimum to the maximum of the observed values of diversity are predicted to affect housing prices. Figure 4 reflects the relatively high implicit price for living in a diverse neighbourhood. Increases in diversity are in particular appreciated if the neighbourhood initially has a very low level of country background diversity.

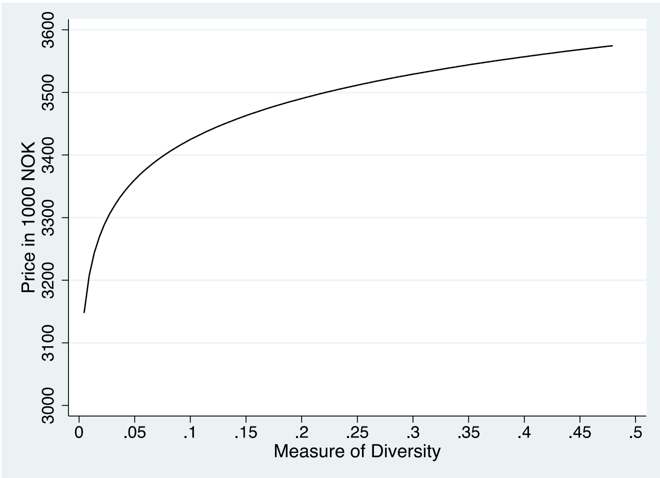
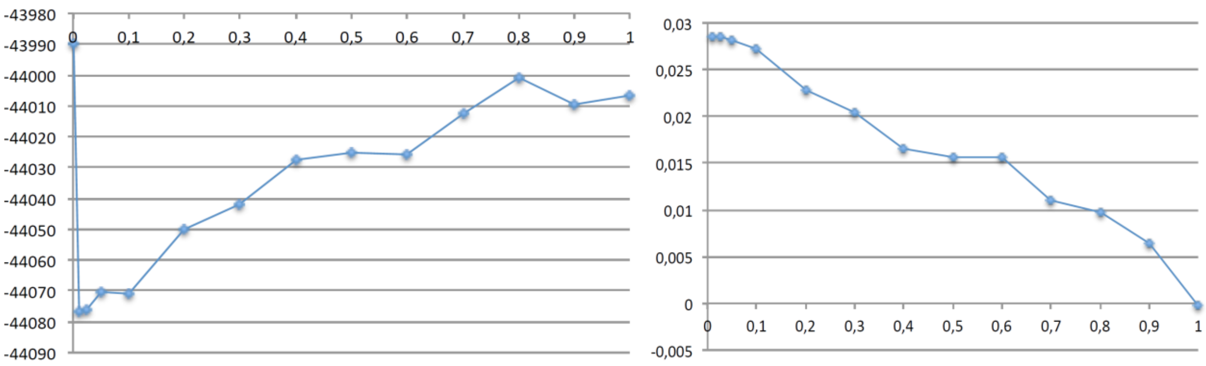


Figure 4: The effect of observed neighbourhood variation in the diversity of country background on the price of a standard house; k = 0.1.

Figure 4 illustrates the importance of using a flexible measure of diversity in relation to housing prices. The threshold for a country background to be classified as properly represented in a neighbourhood, k, is a key parameter. The impact of variations in k on the model performance (DIC) is illustrated in Figure 5a. A value of $k = 0$ means that country background is ignored, while k = 1 corresponds to a thick definition of diversity. We see that $k \leq 1$, gives better model performance than $k = 0$. Hence, diversity (or its absence) matters in relation to housing prices. Moreover, we see that relatively low values of k give a better model performance than values corresponding to thicker definitions of diversity. It is tempting to speculate that the presence of diversity matters more than its thickness. We might paraphrase Ottaviano and Peri (2006): It's nice to have Italian restaurants, French beauty shops, German breweries, Belgian chocolate stores, Russian ballets, Indian tea houses and Thai massages parlours in your neighbourhood but it's not that important that there are many of each of them. A thin diversity means that people from a wide range of countries are represented in a neighbourhood, but only by relatively low proportions of the total local employment.



a) DIC b) β_5

Figure 5: a) Model performance (DIC values) and b) impact of variations in k on how country background diversity of the inhabitants affects the housing prices in a neighbourhood (β_5).

Figure 5b illustrates how the marginal impact on predicted house prices of variations in diversity varies with k. It is clear from Figure 5a that the effect of diversity is better captured by a k-value of 0.1 than by a value of k = 1. The marginal impact of increased diversity approaches zero as k approaches 1. For k = 1, the estimated effect of variations in diversity is not significantly different from 0.

Hence, our results mean that households in the Oslo housing market on average value a diverse, thin, representation of neighbours from a wide range of countries. A natural extension of our analysis could be to study the persistence of thin diversity in a neighbourhood that is to introduce a dynamic perspective to a study of residential location decisions. Does a neighbourhood with a thin diversity of households with respect to land background attract a diverse set of new households? Could this bring a neighbourhood into a state of thick diversity, creating a disamenity over time? By studying residential location decisions of economically diverse in-migrants, McKinnish and White (2011) discuss issues relating to neighbourhoods in transition and the stability of diverse neighborhoods. It would be interesting to conduct a similar approach for diversity defined by land background rather than income, but this is beyond the scope of this paper, and left for future research.

We also experimented with the RDI measure of diversity mentioned in Section 5.1. Once again, country background diversity was estimated to give significantly higher house prices. However, the model performance was poorer than for the models using DIV and the results are not reported here.

The estimated effect of variations in country background diversity is interpreted as a causal relationship. As far as we can see, there is no obvious theoretical reason why this should be a case of reverse causality. At the same time, we do not believe that our approach suffers seriously from omitted relevant information, and we have seen that the parameter estimate related to diversity is very consistent across alternative model formulations.

7. Concluding remarks

Our main goal in this study was to explore how local concentrations and diversity of households with different country background are reflected in housing prices in a neighbourhood. To avoid estimation bias related to the impact of variables representing the segregation of immigrants, we pursued two strategies. First, we controlled for a large set of observed socio-economic and demographic variables characteristics of neighbourhoods. An unusually rich and detailed dataset enabled this strategy. Second, we used a conditional autoregressive spatial model formulation to capture potentially relevant omitted spatially correlated non-observable information of zonal characteristics. The importance of this extensive controlling approach becomes evident when we compare the results that show a weak effect (i.e. the low absolute magnitude of the coefficients of country background shares) in our preferred model with the high bivariate correlation between housing prices and immigrant shares.

Households in the Oslo housing market on average prefer neighbourhoods with a high concentration of native Norwegians, many immigrants from Western countries and, at the same time, a diverse, thin, representation of neighbours from a wide range of countries. We find no evidence that some minorities are considered a kind of nuisance that market actors undertake costly actions to avoid (Yinger, 2016). In other words, on average, we find no marked tendencies that immigrants from certain countries are not wanted as neighbours by the majority of house buyers in the Oslo urban area. The coefficients of some of the non-Western groups are indeed significant and below zero, for example Somalia, Central Asia and Turkey. However, the magnitude of the estimated effects is negligible.

The finding that the local share of native Norwegians has a substantial positive impact on housing prices in a neighbourhood, is consistent with what Cutler et al. (1999) termed decentralized racism. Neighbourhoods with a high concentration of majority members are considered to be attractive in the housing market, ceteris paribus. This very significant effect implies that there is a relatively strong negative impact on housing prices related to the presence of non-natives in a neighbourhood. Following Accetturo et al. (2014, p. 55), this result suggests that increased immigration to a neighbourhood represents a deterioration of local amenities.

Nevertheless, a main result in this paper is that house buyers in the Oslo urban area on average evaluate a diversity of inhabitants with different country background as an attractive amenity of a neighbourhood. This could represent a causal relationship; and a diverse neighbourhood has a positive impact on the local housing prices. Although the empirical literature on this issue is scant, our results do not entirely concord with those by Li (2014) who found that neighbourhoods with diverse minority composition tended to be lower priced compared to neighbourhoods having more homogeneous ethnic compositions. However, Ottaviano, and Peri (2006) found patterns that are more similar to our results.

This study demonstrates that it is important to employ a flexible measure of diversity. Using this approach, we find that house buyers in the Oslo urban area tend to prefer a thin diversity, where many nationalities are represented by a few members rather than many members, relative to the overall local population. Because the topic explored here is empirically under-researched, we believe that our paper contributes to the understanding of segregation patterns, mobility, the associated functioning of housing markets, and urban social sustainability.

The Oslo metropolitan area is one among many European cities that have experienced a growth in its non-native population. As such, our study should also be of interest to a wider audience. Moreover, segregation of major minorities is relatively low in Oslo in comparison to larger cities in other countries.

In interpreting the findings, one should also remember a special feature of both the Norwegian and the Oslo housing market: namely, that homeownership is the dominant tenure. The dominance of home ownership in a city where new construction comprises only a tiny part of the already built housing stock leaves public authorities with limited policy options. Yet, this makes it even more crucial to utilize the available opportunities. To promote diversity, land use regulations and planning practices should not contribute to further homogenization of neighbourhoods. More concretely, public planning should facilitate new construction of terraced and single-family housing in those parts of the city that are dominated by blocks of flats, and smaller housing units in blocks of flats where terraced and single-family housing dominates. Even more critical are the policies that affect utilization of the existing housing stock. Utilizations of the (small) public housing stock should be guided by an urge to stimulate diversity at the local level. The purchase of existing dwellings for new public renting should not be restricted to the least expensive parts of the city. Finally, the Norwegian State Housing Bank runs a large-scale programme providing loans for marginal homeowners. Some groups of immigrants are over-represented in this programme. Diversity at the local level and combating segregation rather than just partial cost efficiency should be an important part of their plan.

It is by no means obvious that the low negative marginal effects of variations in the shares of inhabitants with a non-Norwegian country background are global characteristics of the Oslo housing market. It could just as well be the case that a low level of segregation and low marginal effects are part of a number of possible multiple equilibria. Hence, such an equilibrium has positive properties, especially in terms of social cohesion, and effort should be made to avoid the risk of alternative equilibria with more segregation and higher marginal effects. This risk forms an additional argument for the policy means proposed above.

Finally, despite the steps, we have taken to avoid endogeneity and identification problems, some methodological issues remain to be addressed in future research. This might involve the search for better instruments and/or a clear natural experiment, as well as accounting for distance in modelling spatial autoregressive processes, rather than simply relying on contiguity in the specification of the binary neighbourhood matrix.

References

Accetturo, A, F Manaresi, S Mocetti, and E Olivieri, 2014, “Don’t stand so close to me: The urban impact of immigration”, *Regional Science and Urban Economics*, 45, 45-56.

Alesina, A, and E La Ferrara, 2005, “Ethnic diversity and economic performance”, *Journal of Economic Literature*, 43(3), pp. 762–800.

Amior, M and A Manning, 2018, “The persistence of local joblessness”, *American Economic Review*, 108, 7, 1942-1970.

Anselin, L, 1988, Spatial econometrics: methods and models, London, UK, Kluwer Academic Publishers.

Bakens, J, P Mulder, and P Nijkamp, 2013, “Economic impacts of cultural diversity in the Netherlands: Productivity, utility and sorting”, *Journal of Regional Science*, 53(1), pp. 8–36.

Baum, C F, M E Shaffer, and S Stillman, 2003, “Instrumental variables and GMM: Estimation

and testing", *The Stata Journal*, 3(1), pp. 1-31.
Bayer P., McMillan R., 2008, Distinguishing Racial Preferences in the Housing Market: Theory and Evidence. In: Baranzini A., Ramirez J., Schaerer C., Thalmann P. (eds) Hedonic Methods in Housing Markets. Springer, New York, NY
Bivand, R, S Zhe, L Osland, I S Thorsen, 2017, "A comparison of estimation methods for multilevel models of spatially structured data", Spatial Statistics, 22, Part B, pp. 440–459.
Borjas, G. (1992) "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics* 107 (1): 123-50.
Brueckner JK, Thisse J-F and Zenou Y. (1999) Why is central Paris rich and downtown Detroit poor?: An amenity-based theory. *European Economic Review* 43, pp. 91-107.
Cheshire, P and S Sheppard, 1998, "Estimating the demand for housing, land, and neighbourhood characteristics", *Oxford Bulletin of Economics and Statistics*, 60, pp. 357–382.
Cobb-Clark, D A and M G Sinning, 2011, "Neighbourhood diversity and the appreciation of native- and immigrant-owned homes", *Regional Science and Urban Economics*, 41, pp. 214–226.
Cutler, D M, E L Glaeser, and J L Vigdor, 1999, "The rise and decline of the American ghetto", *Journal of Political Economy*, 107, pp. 455–506.
Delmelle, E C, 2017, "Differentiating pathways of neighbourhood change in 50 US metropolitan areas", *Environment and Planning A*, 49, 2401–2424.
Ellen I G (2000) Sharing America's Neighborhoods: The Prospects for Stable Racial Integration. Cambridge, MA and London: Harvard University Press.
Ellen, I G and K O'Regan, 2010, "Welcome to the neighbourhood: How can regional science contribute to the study of neighbourhoods?", *Journal of Regional Science*, 50, pp. 363–379.
Gelman, A, J. Hwang, and A. Vehtari, 2014, "Understanding predictive information criteria for Bayesian models", *Statistics and Computing*, 24, pp. 997–1016.
Gonzalez, L and F Ortega, 2013, "Immigration and housing booms: Evidence from Spain", *Journal of Regional Science*, 53, pp. 37–59.
Gschlößl, S and C Czado, 2007, "Spatial modelling of claim frequency and claim size in non-life insurance", Scandinavian Actuarial Journal, 3, pp. 202–225.
Guerrieri, V, D Hartley, and E Hurst, 2013, "Endogenous gentrification and housing price dynamics", *Journal of Public Economics*, 100, pp. 45–60.
Hwang, J and R J Sampson, 2014, "Divergent pathways of gentrification: Racial inequality and the social order of renewal in Chicago Neighborhoods", *American Sociological Review*, Vol. 79 (4), 726-751.

Korn, R, E Korn, and G Kroisand, 2010, Monte Carlo methods and models in finance and insurance, Financial Mathematics Series, Chapman and Hall, 1st edition.

Li, Q, 2014, “Ethnic diversity and neighbourhood house prices”, *Regional Science and Urban Economics*, 48, pp. 21–38.

McCann, P, 2013, Modern Urban and Regional Economics, Oxford University Press.

McKinnish, T and T Kirk White, 2011, “Who moves to mixed-income neighborhoods?”, *Regional Science and Urban Economics,* 41, 187-195.

Moye, R, 2014, “Neighbourhood racial-ethnic change and home value appreciation: Evidence from Philadelphia”, *Urban Geography*, 35, pp. 236–263.

Musterd, S, 2005, “Social and ethnic segregation in Europe: Levels, causes and effects,” *Journal of Urban Affairs*, 27, pp. 331–348.

Nordvik, V and L Osland, 2017, “Putting a price on your neighbour.” *Journal of Housing and the Built Environment*, 32, pp. 157–175.

Oates, WE (1969), “The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis”, *The Journal of Political Economy*, 77(6), pp. 957-971.

OECD/European Union (2015), Indicators of Immigrant Integration 2015: Settling In, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264234024-en

Osland, L and I Thorsen, 2008, “Effects on housing prices of urban attraction and labor market accessibility”, *Environment and Planning A*, 40, pp. 2490–2509.

Osland, L, I Sandvig Thorsen, and I Thorsen, 2016, “Accounting for spatial heterogeneities in housing market studies”, *Journal of Regional Science*, 56, pp. 895–920.

Ottaviano, GIP and G Peri, 2006, “The economic value of cultural diversity: Evidence from US cities”, *Journal of Economic Geography*, 6, 9–44.

Reed, W R, 2015, “On the practice of lagging variables to avoid simultaneity,” *Oxford Bulletin of Economics and Statistics*, 77, pp. 897–905.

Saiz, A, 2007, “Immigration and housing rents in American cities”, *Journal of Urban Economics*, 61, 345-371.

Saiz, A and S Wachter, 2011, “Immigration and the neighbourhood”, *American Economic Journal: Economic Policy*, 3, pp. 169–188.

Spiegelhalter, D, A Thomas, N Best, and D Lunn, 2003, WinBUGS user manual.

Tiebout, C. (1956), "A Pure Theory of Local Expenditures", *Journal of Political Economy*, 64 (5): 416–424, *doi*:*10.1086/25783*

Vang, Z M, 2010, “Housing supply and residential segregation in Ireland”, *Urban Studies*, 47,

pp. 2983–3012.
Wachter, S, 2017, ”Immigration and segregation”, Cato Journal, Vol. 37, No.3, 495-501.
Wessel, T, 2016, “Economic segregation in Oslo: Polarization as a contingent outcome”. Ch. 6 in T Tammaru, S Musterd, M van Ham and S Marcinczak (eds.), Socio-Economic Segregation in European Capital Cities: East Meets West, Routledge. ISBN 9781138794931.
Wessel T and Nordvik V. (2018) Mixed neighbourhoods and native out-mobility in the Oslo region: The importance of parenthood. *Urban Studies*: https://doi.org/10.1177/0042098018768452
Yinger, J, 1976, Racial prejudice and racial residential segregation in an urban model, *Journal of Urban Economics*, 3, pp. 383-396.
Yinger, J, 2016, “Hedonic estimates of neighbourhood ethnic preferences”, *Public Finance Review*, 44, pp. 22–51.

Appendix:

A list of covariates in the linear predictor and results on location characteristics, dwelling attributes and year dummies.

Table 3: Specific covariates.

	Dwelling attributes
DETACHDUM=	1 if the observation is a detached house, 0 if it is not (dummy)
TERRDUM=	1 if the observation is a terraced house, 0 if it is not (dummy)
SEMI-DETDUM=	1 if the observation is a semi-detached house, 0 if it is not (dummy)
LIVEAREA(DET)=	the living area of a detached house, measured in square meters
LIVEAREA(SMA)=	the living area of a small house, measured in square meters
LIVEAREA(APA)=	the living area of an apartment, measured in square meters
CONYEAR45=	1 if the house was constructed in 1940-50, 0 if not (dummy)
CONYEAR60=	1 if the house was constructed in 1951-70, 0 if not (dummy)
CONYEAR75=	1 if the house was constructed in 1971-80, 0 if not (dummy)
CONYEAR85=	1 if the house was constructed in 1981-90, 0 if not (dummy)
CONYEAR95=	1 if the house was constructed in 1991-00, 0 if not (dummy)
CONYEAR03=	1 if the house was constructed in 2001-06, 0 if not (dummy)
CONYEARREC=	1 if the house was constructed in 2007-, 0 if not (dummy)
COOPDUM=	1 if the building is a part of a housing cooperative, 0 if not (dummy)
SHAREDUM=	1 if the building is a condominium, 0 if not (dummy)
	Yeardummies
YEARDUM10=	1 if the house/apartment was sold in 2010, 0 if not (dummy)
YEARDUM11=	1 if the house/apartment was sold in 2011, 0 if not (dummy)
YEARDUM12=	1 if the house/apartment was sold in 2012, 0 if not (dummy)
	Location
TIMECBDW=	the travelling time in minutes from the Oslo city center, to the west
TIMECBDN=	the travelling time in minutes from the Oslo city center, to the north
TIMECBDS=	the travelling time in minutes from the Oslo city center, to the south
ACCESSIBILITY=	labour market accessibility, defined by a standard Hansen type of measure
	Economic and sociodemographic characteristics
AGE1019=	the proportion of the population in a zone in the age group of 10-19
AGE2030=	the proportion of the population in a zone in the age group of 20-30
AGE3040=	the proportion of the population in a zone in the age group of 30-40
AGE4050=	the proportion of the population in a zone in the age group of 40-50
AGE5060=	the proportion of the population in a zone in the age group of 50-60
AGE6070=	the proportion of the population in a zone in the age group of 60-70
AGE70+=	the proportion of the population in a zone in the age group of 70+
RDIAGE=	a measure of the diversity of age groups in a zone
5OUTMOVE=	the proportion of households with children under 5 years old that has moved out of the zone
OUTMOVE=	the proportion of households that has moved out of the zone
HIGHEDU=	the proportion of the zonal population in the age group 30-49 with a bachelor's and/or a master's degree
LOWEDU=	the proportion of the zonal population in the age group 30-49 with no more than secundary education
INC(0711)MED=	the average income in the period 2007-2011 of the median male income earner in the zone
INC(0711)SD=	the standard deviation of the average 2007-2011 incomes for male income earners in the zone
INCGT120=	the proportion of the population in the zone with income higher than 120000 NOK
SOCSECBEN=	the proportion of the population in the zone receiving Social Security benefits
NATSECBEN=	the proportion of the population in the zone receiving National Security benefits
CASHCARE=	the proportion of the households in a zone receiving cashcare for small children
DETPROP=	the proportion of detached houses in the zone
APAPROP=	the proportion of apartments in the zone
BLOCKPROP=	the proportion of blocks of flats in the zone
SMALLPROP=	the proportion of small houses in the zone
	Country Background
NORWEGIAN=	the proportion of the population in the zone with background from Norway
POLAND=	the proportion of the population in the zone with background from Poland
RUSSIA=	the proportion of the population in the zone with background from Riussia
OCEANIA=	the proportion of the population in the zone with background from Oceania
SOMALIA=	the proportion of the population in the zone with background from Somalia
TURKEY=	the proportion of the population in the zone with background from Turkey
CENTRALASIA=	the proportion of the population in the zone with background from Central Asia
DEVELOPING=	the proportion of the population in the zone with background from developing countries (except Somalia)
WESTERN=	the proportion of the population in the zone with background from western countries (except Norway, Poland, and Russia)
NON-WESTERN=	the proportion of the population in the zone with background from non-western countries
DIV_k =	the diversity of the population in a zone, w.r.t. country background, for a specific value of k

Table 4: Results on location characteristics, dwelling attributes and yeardummies.

	M1(A)	M2(A)	M1(DA)	M2(DA)	M3(DA)
Constant	9.3***	9.129***	9.078***	9.409***	9.335***
	(0.1711)	(0.1685)	(0.1728)	(0.1706)	(0.1704)
TIMECBDW	2.379E-4	1.579E-4	0.003591***	-0.01797*	-0.01688*
	(0.008677)	(0.008676)	(0.008859)	(0.008948)	(0.008959)
TIMECBDN	-0.06628***	-0.06332***	-0.05806***	-0.07325***	-0.07162***
	(0.01285)	(0.01277)	(0.01153)	(0.01192)	(0.0119)
TIMECBDS	-0.01525	-0.0177	-0.006913	-0.03065***	-0.03024***
	(0.01145)	(0.01142)	(0.01112)	(0.01085)	(0.01086)
ACCESSIBILITY	-0.01086*	-0.01138**	-0.006844	-0.0178***	-0.01663***
	(0.005735)	(0.005725)	(0.005659)	(0.005667)	(0.005664)
LIVAREA(DET)	0.6025***	0.6016***	0.6036***	0.6008***	0.6012***
	(0.005735)	(0.005734)	(0.005792)	(0.005763)	(0.005765)
LIVAREA(SMA)	0.6438***	0.6425***	0.6459***	0.6424***	0.6418***
	(0.00643)	(0.006429)	(0.006412)	(0.006406)	(0.006408)
LIVAREA(APA)	0.7703***	0.7709***	0.7704***	0.7706***	0.7704***
	(0.002004)	(0.002004)	(0.001988)	(0.001979)	(0.001979)
CONYEAR45	0.001938	0.001586	0.00115	9.527E-4	0.001329
	(0.004505)	(0.004506)	(0.00452)	(0.004516)	(0.004518)
CONYEAR60	-0.02653***	-0.02516***	-0.02447***	-0.02653***	-0.02621***
	(0.002463)	(0.00246)	(0.002459)	(0.002435)	(0.002437)
CONYEAR75	-0.02153***	-0.02284***	-0.02305***	-0.02421***	-0.02282***
	(0.00321)	(0.003204)	(0.003205)	(0.003219)	(0.003225)
CONYEAR85	-0.003363	-0.002855	-0.005137*	-0.001855	-6.277E-4
	(0.002969)	(0.00297)	(0.002965)	(0.00296)	(0.002959)
CONYEAR95	0.04665***	0.04671***	0.04486***	0.04672***	0.04835***
	(0.00318)	(0.003179)	(0.00322)	(0.00322)	(0.003219)
CONYEAR03	0.1251***	0.125***	0.1231***	0.1256***	0.1276***
	(0.00293)	(0.002929)	(0.002914)	(0.002912)	(0.00291)
CONYEARREC	0.1292***	0.1289***	0.1273***	0.1277***	0.13***
	(0.002717)	(0.002714)	(0.00266)	(0.002704)	(0.002708)
COOPDUM	-0.00111	-7.775E-4	-0.001748	-2.448E-4	-6.531E-4
	(0.001798)	(0.001794)	(0.00179)	(0.001796)	(0.001796)
SHAREDUM	-0.006986**	-0.007822**	-0.007316**	-0.006267*	-0.006658*
	(0.003469)	(0.003468)	(0.003496)	(0.003501)	(0.003503)
DETACHDUM	0.9064***	0.914***	0.902***	0.9164***	0.913***
	(0.03078)	(0.03078)	(0.0311)	(0.03093)	(0.03094)
TERRDUM	0.6366***	0.6472***	0.6275***	0.6455***	0.6471***
	(0.03181)	(0.03181)	(0.03179)	(0.03144)	(0.03145)
SEMI-DETDUM	0.6461***	0.6559***	0.6365***	0.6542***	0.6555***
	(0.03284)	(0.03284)	(0.03288)	(0.03253)	(0.03254)
YEARDUM10	0.08283***	0.08282***	0.08278***	0.08272***	0.08272***
	(0.001729)	(0.001729)	(0.001715)	(0.00171)	(0.00171)
YEARDUM11	0.1825***	0.1824***	0.1825***	0.1823***	0.1823***
	(0.001696)	(0.001695)	(0.001669)	(0.001691)	(0.001692)
YEARDUM12	0.2505***	0.2505***	0.2503***	0.2507***	0.2507***
	(0.001877)	(0.001877)	(0.001883)	(0.001884)	(0.001885)

Note: * significant at the 0.1 level, ** significant at the 0.05 level, *** significant at the 0.01 level; robust standard errors in parentheses.