Putting a price on your neighbour

Viggo Nordvik

NOVA, Oslo and Akershus University College, viggo.nordvik@nova.hioa.no

Liv Osland

Haugesund and Stord University College,

liv.osland@hsh.no

Abstract

Neighbourhood population composition affects the willingness to pay for housing units. This paper utilises a large and rich data set and hedonic regression techniques to disentangle the effect that neighbourhood affluence and presence of inhabitants with an immigrant background have on home prices. Furthermore, we specify an empirical model in a way that also enable us to test for the effect of diversity, both in terms of income levels and of the composition of the immigrant population of a neighbourhood. The hedonic model can be viewed as a variety of an amenity interpretation of the population composition of a neighbourhood. Estimation of effects of population composition is not straightforward as there is good reason to believe that population composition is both endogenously determined together with house prices and that area level omitted variables could bias estimates. This is addressed by lagging the composition measures and by formulating two different models that address these difficulties in different ways. We estimated one random effects model that instruments within neighbourhood variation in population composition and one fixed effects model that control for omitted variables. We find that coefficient estimates are robust across these specifications.

Keywords: Home prices, neighbourhood characteristics, diversity, Urban sorting, Oslo

1. Introductory remarks

If you read newspapers, and especially the comment fields under articles dealing with neighbourhood change and mix, you most certainly get the impression that many people care about the socio-economic and ethnic composition of their neighbourhoods. The same impression holds true if you talk to local inhabitants, or if you stay at home reading highly ranked academic journals as e.g. the American Economic Review or Journal of Political Economy (Hoff and Sen (2005), Sethi and Somanathan (2004) and Wigren (1987)). People care for social and ethnic mix simply because they care (i.e. preferences), because others (are believed to) care or because they believe that certain mixes yield undesirable outcomes – while other mixes contribute to more desirable outcomes. When people have preferences towards the socio-economic composition of neighbourhoods, this will be reflected in their residential locational choices and, hence, housing markets.

Parallel to this, recent developments have shown that a relative large entry of immigrants to urban areas mainly from developing countries, has contributed to significant growths in population. By way of example, the immigrant share of the population in Norway in 1990 was around 4 %. In 2014 the per cent of immigrants was almost 15. The growth has been particular strong in the Oslo area, where 31 % of the population are immigrants or have immigrant parents (Statistics Norway). This has contributed to an increase in the demand for housing, but more importantly, it has made many residential areas more ethnically diverse than they used to be. To some extent this may augment or change the variation in real estate prices within an urban landscape. In addition to traditional important spatial structure characteristics like access to labour markets, housing prices may have become more profoundly related to socioeconomic composition or structure of neighbourhoods.

House price differentials between immigrant-dense and not so dense areas are not much studied in a Norwegian, or even a European context. The US have a much longer history with mixed societies and the topic has popped up more or less regularly, at least since Bailey (1959) and Bailey (1966). As we test for tipping points, our study also relates to the emerging literature on tipping in majority shares as a neighbourhood becomes more minority dense, Card et al. (2008). Hence, even though our topic *Putting a price on your neighbour* is somewhat uncomfortable, we will claim that understanding the interplay between population composition and prices of owner-occupied housing is crucial in understanding neighbourhood change. The tragic history of racial tension and urban decline in Detroit and the drivers of change there as described by Galster (2012), can serve as an illustration. His descriptions of the (sometimes violent) actions taken by some people to protect the 'value of their property' from the threat of black in-migration most certainly warrants a study of possible tipping of house prices in a situation with a growing non-native population in many of our neighbourhoods.

This paper is an empirical investigation of how different characteristics of neighbourhood mix affect prices of owner-occupied housing in the wider Oslo metropolitan region. This study area is very proper for the analysis to be formed, given that the vast majority (around 80%) of households own their houses. This is also the case for immigrants, and 63 % of this group owns the dwelling they live in Blom and Henriksen (2008). We do not use information on ethnicity or citizenship when defining who belongs to a minority. Rather we use what Statistics Norway term land background. It is a classification based on the country of birth of both a person and of her parents. As such it is a kind of country of origin measure.

Empirically we use the hedonic method of Rosen (1974), which is a revealed preferences approach, based on market outcomes. To be more specific, the approach that we use originates from Yinger (1976) in which neighbourhood composition is treated as an amenity that capitalizes into house prices. In order to identify the implicit prices of the studied neighbourhood characteristics we need to compare the prices of equivalent houses. There is, hence, a need to control for a battery of other observable and unobservable factors. According to Bayer and McMillan (2008) unobservables may be of a particular concern when the focus is on aspects related to racial preferences in relation to housing prices. They argue that there may be systematic correlation between unobserved neighbourhood quality and the proportion of immigrants. We address this topic in part by using a fixed effect approach and in part by using a random effects approach that enables us to distinguish between within and between extended neighbourhood variations in the population composition.

Our analyses can be seen as a confrontation between a rich and close to full coverage empirical data set on cross-sectional variation in home prices in the Oslo region, Norway, and two different strands of literature. On the one hand, we have the white flight/avoidance literature that treats (some) minorities as a nuisance that some agents undertake costly action in order to avoid, Yinger (2014), Cutler et al. (1999) and Bailey (1966). On the other hand, we have the Florida-inspired literature that focus on productivity gains from a diverse population and a diverse set of consumption possibilities arising out of a culturally diverse population, Ottaviano and Peri (2006), Nathan (2012) and Bakens et al. (2013).

The larger part of the paper is about correlations that can be interpreted within the frame of a set of hypotheses derived from prior empirical and theoretical studies. We demonstrate e.g. that the share of neighbours somewhat dissimilar to the majority pushes home prices downwards and that the median neighbourhood income commands a positive price response. The evidence on the effects of diversity is less clear-cut.

In the next section we briefly present some theoretical arguments for the hypotheses tested in the empirical analysis. Secondly, we give a short overview of findings in the hedonic house price literature that are relevant for our research. Then we discuss how to specify an empirical model. Next, we give some empirical background before we proceed with the empirical results on price variation in the Oslo region. Here we do both test for discontinuities in the price process and undertake some sensitivity analyses. Finally, we provide concluding remarks.

2. Preferences, sorting and home prices

2.1 Preferences over neighbourhood mix

Throughout the paper, we take as a basic premise that people prefer some (types of) neighbours to others. Starting from this, we test whether and to what extent such preferences feed into house prices. In the literature there exists direct attempts to test for such preference, e.g. the card-experiments where respondents rank hypothetical neighbourhoods with different mixes in a stated preference setting, see Farley et al. (1978) and Charles (2003). There have been proposed different answers to the question of why individuals care for who their neighbours are.

A simple answer is that people care because they care. In other words, some hold a kind of Beckerian preferences over the land background, or even ethnicity, of their neighbours. Other answers that has been proposed in the literature is that cultural distance reduces the efficiency of communication, and consequently increases the burden of misunderstandings, and that enforcement of (both implicit and explicit) contracts is more cumbersome across cultural 'borders' (Alesina and La Ferrara (2005)). The neighbourhood is an arena of random meeting with people, Li (2014) hypothesise that the benefits this kind of informal network provide is more valuable if it facilitates random meetings with your 'own' kind.

It has been demonstrated that even quite weak preferences for not being a minority can yield strongly segregated spatial outcomes (Schelling (1971)). Furthermore, under not very unrealistic conditions a complete segregation can arise even if everyone prefers to be a minority to being a majority in the neighbourhood. This result can arise if you prefer to be part of a dominant majority to being part of a very small minority (Zhang (2011)).[1] Voting with the feet in this manner can yield spill-overs into house prices. This can give rise to an alternative reason to care for, e.g. country background, composition of neighbourhoods. I.e. that people have preferences for neighbourhood composition because they believe (or even fears) that others have such preferences. Buying a home is a large investment, and if you fear that the area will be less attractive to others in the future, you will reduce your maximum bid for a property – simply because most of us want to protect the value of our investments.

That leads us to the next type of arguments for preferring some mixes to others. You may rightly or wrongly, believe that some mixes yield favourable outcomes while others do not. This can give rise to self-enforcing effects where the housing units preferred by e.g. the rich command an extra price because there is a view that rich people invest more in their neighbourhoods, by taking care of properties and by having an ability to take action when needed. Under the reasonable assumption that the presence of affluent people in a neighbourhood is a normal good, this can give rise to social multipliers where rich people cluster together with rich people. *An interplay of within-community externalities and market forces can lead to cities that are segregated by tenure and income – with the rich living with the rich in homeowner communities with well-functioning civic environments, and the poor living with the poor in dysfunctional renter communities* (Hoff and Sen (2005): pp. 1167-68). In essence, this is a social capital argument and related impacts on home prices can be viewed as a capitalising of neighbourhood effects.

An analysis somewhat similar to that of Hoff and Sen (2005) has been provided by Sethi and Somanathan (2004). They show how the interplay between preferences for the affluence of neighbours, their ethnicity and an income distribution where the majority dominates (not

[1] One important implication of this is that you cannot derive strong conclusions on micro-level attitudes and preferences from observations of macro-level observations.

necessarily strictly) the minority could create segregated outcomes. In this segregated outcomes, the more affluent members of the outbid minority members in neighbourhoods with a high share of majority members. We note that Sethi and Somanathan's study demonstrate the possibility of multiple equilibria, and how social policy initiatives that enable some families to settle down in another type of neighbourhood than they would have done otherwise, can help moving the local distribution into a more even spread of minorities and majority and reduce price differentials between areas.

3. Housing prices and neighbourhood mix: Some results from the hedonic house price literature

Neighbourhoods may be defined as "the bundle of spatially based attributes associated with clusters of residences, sometimes in conjunction with other land uses" (Galster (2003)). When present, perceived and valued, these spatial attributes capitalize into housing prices. In this way the overall value of neighbourhoods may be determined by inter alia, accessibility to labour markets (Osland and Thorsen (2008)), access to more or less clean environment (Yinger (2014)), or shared social demographic features like the risk for crime (Dubin and Goodman (1982)), and ethnic diversity or homogeneity (Li, 2014)).

An overview of results on neighbourhood characteristics in the hedonic house price literature is found in Chin and Chau (2003). A summary of the literature of racial discrimination and prejudice in the US, and how these features capitalize into the price of housing is provided by Zabel (2008). He finds that the bulk of studies are based on data from the 1970's, and there are relative few recent studies; mainly due to the lack of relevant data. Notable exceptions are DeSilva et al. (2012) and Li (2014), both based on US data. Li (2014) study the impact of ethnic diversity on housing prices by using a range of different panel data methods to rule out the possibility of coefficient bias. One robust result is that neighbourhoods with relatively homogeneous minority populations are higher priced than neighbourhoods with more diverse minority compositions.

DeSilva et al. (2012) emphasize that the empirical literature on how race and ethnicity impact on housing markets has primarily studied two issues: Whether minorities pay more than whites for an equivalent house; and whether the presence of minorities have a negative impact on housing prices (page 242). The older literature does not agree on the answer to these questions (DeSilva, 2012). More recent papers also show varying results, dependent inter alia on model specification (Baumont (2009). It is argued that biased results are probable because of misspecifications, unobservable and unmeasurable features and sorting (Bayer and McMillan, (2008)). One frequently mentioned problem is the strong correlation between ethnic compositions of neighbourhoods and other physical or social demographic neighbourhood characteristics, many of which are not included in the model specifications.

In spite of the long term interest in the impacts of race, ethnicity diversity and other social characteristics, it is probably correct to claim that this is currently an understudied area in housing market studies (Visser et al. (2008); Herath and Maier (2010)), and the results differ substantially mainly due to data and endogeneity issues. Moreover, most papers are based on US data, where the degree of racial or ethnical transitions and level of segregation of neighbourhoods are more prevalent, have a far longer history and a different structure (Zabel, (2008); Moye (2014)). The results found in these studies, may, hence be less relevant in a European or Nordic setting.

To our knowledge, Wigren (1987) is the only Scandinavian study relevant to this analysis. He aims at finding the determinants of owner-occupied housing in Sweden. He uses the share of inhabitants in a neighbourhood voting for the Conservatives as an indicator of socioeconomic status. As hypothesised, this variable significantly adds a value to a housing unit. Arguably, one could interpret this as a kind of social capital effect; and may capture much of the same as we do with our measure of neighbourhood income. Another relevant study is from Reading and Darlington (UK); here Cheshire and Sheppard (1998) find a negative price elasticity of the share of African-Caribbean population in a ward with an absolute value above 1.

Hence, there should be room for an analysis of the interdependencies between population composition and house price variation in a European setting.

4. Specification of our empirical model

Specification of an empirical model is (almost) always constrained by the type of data we have access to. We will here not start out with a detailed description of specification in a perfect case and discussion of how to cope with the fact that data deviate from our conception of the perfect data to test our hypotheses. Still we will point to facts that limits our choices of specification. Firstly, we are not able to identify repeat sales and we do not have a long time-series, hence a difference-in-difference design is not feasible. Second, as we do not have access to any information about buyers we are not able to identify any differences in preferences (or marginal willingness to pay) between groups. Our estimates should consequently be interpreted as characteristics of the envelope of marginal bid functions of the market participants. Roughly formulated, the estimated coefficients capture the marginal willingness to pay of the 'market'.

Our econometric model should be a frame for testing hypotheses on how population composition in the neighbourhood affects prices. If population composition is independent of other explanatory variables one could simply fit an OLS model to the data in order to obtain estimates of the coefficients of the variables describing population composition. Furthermore, if there are omitted variables at the neighbourhood level that do not co-vary systematically with the population composition variables we primarily are interested in, we could have increased the efficiency of the estimates by including neighbourhood random intercepts.

Unfortunately, there exist sound theoretical arguments and empirical evidence that the conditions for an OLS or a straightforward random effects model to be suitable is violated. In order to estimate the marginal effects of variations in the neighbourhood population composition and moments of the local income distribution we will test out two empirical strategies. First, we utilise an idea described by Zabel (2008). That is, we utilise the fact that we have census tracts nested within clusters of tracts that according to Oslo municipality form natural neighbourhood, and estimates a fixed cluster-of-tracts (hereafter termed: extended neighbourhood or simply neighbourhoods) effect model. Second, we follow Rabe-Hesketh and Skrondal (2008) and formulate an instrumental variable that (test for and) removes the correlation between an extended neighbourhood level random effect and our measures of population composition. Both of these two approaches distinguish between a kind of within and between extended neighbourhood variation and both of them can be interpreted as special cases of the expression.

$$(1)\quad lnP_{ij} = X_{ij}\theta + \left(Z_j\mu + \overline{Z_{j*}}\gamma + \alpha_{j*}\right) + \ \beta_j\omega + \epsilon_{ij}$$

lnP_{ij} is the log of the price of house i, at location j, at a given point in time. The vector of explanatory variables are divided into different sets: X_i is characteristics specific to a housing unit i, Z_j measure different dimensions of the population composition at the tract level, $\overline{Z_{j*}}$ is the average value of the population composition measures within the extended neighbourhood and β_j is other tract level characteristics. Note that we use subscript j to identify census tracts while j* is used to identify clusters of tracts. The prime parameter vector estimated in the model is μ, while θ, γ and ω are estimated coefficients of different controls. The individual level residual ϵ_{ij} is assumed to be standard normally distributed with zero expectation and constant variance.

Different ways of treating the cluster-of-tracts effect α_{j*} and constraints on coefficients yields the two models we utilise in the reminder of the paper. For short, we will term the two models the augmented random effects model (aRE) and the fixed effects model (FE).

<u>The augmented random effects model</u>:

This model treats the α_{j*} as a cluster-of-tracts-level residual: $\alpha_{j*} \sim N(0, \varphi)$. A problem is that the random effect model does not solve the problem of potential bias in the estimated coefficients of the population composition variables due to omitted neighbourhood characteristics. However, a RE-specification where the within-cluster-of-tracts averages of the population composition measures is included, remedies the potential bias due to correlation between α_{j*} the coefficients of interest, Rabe-Hesketh and Skrondal (2008). Equation (1) is (with one minor exception) equivalent to:

$$(2) \quad lnP_{ij} = X_{ij}\theta + \left((Z_j - \overline{Z_{J*}})\mu + \overline{Z_{J*}}\gamma + \alpha_{j*}\right) + \beta_j\omega + \epsilon_{ij}$$

The term $(Z_j - \overline{Z_{J*}})$ is obviously not correlated with α_{j*}, as the mean centring by design removes any such correlation, this is the reason why Rabe-Hesketh and Skrondal (2008) term this a instrumental variable approach. The effect of this is that μ is an unbiased estimator of the within-cluster-of-tracts dependency between population composition and housing prices. Direct estimation of (1) and (2) give identical estimates of both coefficients and standard errors, except for the estimated γ. Using (1) γ will be an estimate of the difference between the within and between clusters-of-tracts-variation in the effect on housing prices[2], while (2) gives an estimate of the between clusters effect.

<u>The fixed effects model (FE)</u>:

The FE-model is estimated by using dummy variables for each of the 173 clusters-of-tracts. Obviously, this consumes a larger number of degrees of freedom; this is, however, not any

[2] This again provides us with a useful test of whether the between and within effects are similar. If they are similar, they are not biased because of omitted variables at the neighbourhood level.

major concern when fitting the model on close to 100,000 observations. This model constrains the γ's to equal zero; or to say it in another way, the γ's is not identified under the FE-approach as the between variation is absorbed into the set of clusters-of-tracts dummies.

The major advantage of using the fixed effects approach is that t*hese will capture an unobserved [..] characteristics that are correlated with the percent non-white and hence will alleviate any associated bias*, (Zabel, 2008) p. 192. Hence, the FE-approach will improve our estimates of the μ-coefficients.

Housing attributes

In our regression models, we include a continuously varying measure of size and dummies for year of construction and four different house types. The size measure is interacted with the house type dummies, allowing for different prices of floor space in different house types. For housing units located in blocks of flats we also include an indicator of floor at which the unit is situated. Admittedly, this set of explanatories is rather parsimonious. One can note that both Bakens et al. (2013) and Ottaviano and Peri (2006) rely primarily on aggregate prices per square meter when they study country background diversity and cross city variation in home prices. Hence, even though our controls for housing attributes are coarse we do control for some of the most important sources of unit specific variation in both price levels and prices per square meter.

Regional accessibility variables

Modelling of spatial interactions follows Osland and Thorsen (2008). In this paper, regional accessibility is captured by two variables: distance to the central business district (CBD) and access to labour markets. These two variables capture the major part of the intraregional spatial variation in housing prices (Osland and Thorsen, 2013). The estimated coefficients are interpreted as the effect of "urban attraction". The variable captures households' evaluation of the various urban amenities found in the CBD of the metropolitan area. In contrast with Osland and Thorsen (2008) the urban attraction effect is not assumed to be isotropic or ring-like. We hypothesise that the influence of travel time depends of the direction in which you travel. Consequently, we interact the travel time variable with dummies describing whether the housing unit is locate to the North, the South or the West of the City hall. Hence, we measure the effect of travel time on house prices along three rays starting approximately at the Oslo City centre. Where distance should be measured from is determined by the data. We choose the postal zone as the CBD that maximizes the explanatory power of the model (R^2).

Access to labour market is important for housing prices. Following Osland and Thorsen (2008) we apply a gravity-based accessibility measure (Hansen, 1959). This measure captures the potential for interaction between an origin *j* and all the postal delivery zones *k* in the study area, so that the sum of jobs in each destination *j*, is weighted by a distance decay function as follows:

$$ACC_j = \sum_{k=1}^{495} E_k \exp(\sigma d_{jk})$$

In the above expression E_k represents employment in postal delivery zone k, d_{jk} is driving time between zone *j* and *k* measured in minutes, while σ is a parameter estimated by using a Maximum Likelihood estimator. The estimated value of the distance deterrence parameter σ is imputed into the above expression, and ACC_j can subsequently be used as a variable in the regressions to follow.

Local neighbourhood attributes

Local neighbourhood attributes are those observed characteristics of the neighbourhood that are part of the vector β_j. We distinguish between neighbourhood attributes that primarily work as controls in our analysis contained in β_j, and the variables of prime interest: the composition of the population in the neighbourhood contained in Z_j. The set of neighbourhood controls include e.g. composition of the housing stock and share of public housing units in the neighbourhood. All these variables are measured at the census tract level. In the FE-model, α_{j*} control for unobserved neighbourhood attributes (shared) at the cluster-of-tracts level.

Neighbourhood composition

Our prime interest in this paper lays in the estimations of the coefficients μ of the population composition covariates Z_j that we describe in this section. As discussed earlier, there are arguments related to social capital and agency to hold preferences over the affluence of neighbours (Hoff and Sen, 2005). This is also a characteristic of the equilibria in the sorting model of Sethi and Somanathan (2004). Our model should therefore include a measure of the income level of the neighbourhood, facilitating tests of the role of income in the sorting process. I.e. we should test whether high-income neighbours is an asset that demanders in the housing market have a willingness to pay for. We have chosen to measure the income levels in the neighbourhoods as the log of the median income among males 35 to 60 years of age.

The empirical formulation should also be a frame for test of whether a varied population in terms of income (conditional on the median income) is considered a positively valued asset in a neighbourhood. Income diversity (or spread) can be measured in a large number of ways. Here a simple approach is chosen. We use the standard deviation in the income distribution of the middle-aged male income. As opposed to the median, the standard deviation is sensitive to presence of some extreme outliers – and we get some extreme values. To avoid that these extreme outliers influence the results to much, the values of the standard deviation is truncated at 500 (thousands NOK). This corresponds to the 99-percentile in the distribution of neighbourhood median incomes. The truncation is done in order to avoid letting the extreme income observations of some individuals in some neighbourhoods influence the estimation results to strongly. As a test we include a dummy equal to unity for neighbourhoods (here tracts) where the SD-measure has been truncated, and zero elsewhere.

The next question we address is whether there exist preferences of the country background of neighbours that feed directly into prices of housing units. One particular aspect we would like to draw attention to is the fact that it is not correct to describe the Oslo population as consisting of one large majority and one minority. Rather we have one large majority, Native Norwegians, and a multitude of minorities. In 2006, 72.2 percent of the population in Greater Oslo had a Native Norwegian country background. The five largest country background groupings[3] together comprised 15.1 percent of the population, Non-Norwegian Nordic country background is the largest group with a share of 5.1 percent. Hence, minorities taken together comprise a not insignificant share of the population, but no single group are very large. This is in contrast with previous studies, which mainly have focused on the black and white differences (DeSilva et al, (2012), page 246).

We hypothesise that the main channels for causal effects from the composition of the population according to country background to cross-sectional variation in home prices goes through cultural distance, recognisability and/or contribution to consumption diversity. In order to capture this, we chose to use the aggregate share of inhabitants from Africa and Asia (AA) measured at the census tract, as part of our set of explanatory variables.

One could explore whether agents hold preferences over minorities that differ in intensity and character between groups, and consequently that the presence of different groups correlate

[3] Here we simply use the groupings of country background into 22 categories provided by Statistics Norway. The five largest groups in 2006 was The Nordic group, Other Western Europeans, Other Asians, Pakistanis and Sub-Saharan Africans (not including Somalia).

differently with house prices. This one could do by including the shares of e.g. Pakistanis, Swedes and Somalis in the empirical model. We will not pursue this line of research here.

Next we prepare our model for test of whether and eventually how, the diversity of the AA-population feeds into prices of housing units. Primarily we aim to test whether a certain level of an AA-share depends on whether the AA-population is diversely composed or is dominated by one single group. This can be done in different ways. One can use the Simpson diversity index as Cutler et al. (1999) and Alesina and La Ferrara (2005) do. The Simpson index has a straightforward interpretation. It is the probability that two randomly chosen immigrants in a neighbourhood should differ in country background. We have[4] chosen the use the share of the single largest group within the AA-population, relative to the size of the AA-population, as an explanatory.

4.6 Some other issues on specification

Our estimated price regressions should be interpreted as an investigation of whether the correlations between prices and our explanatories are consistent with our prior expectations based on the discussion earlier on in the paper and existing studies. Still, we want to avoid, as far as possible, problems of simultaneity between prices, minority shares and the income measures used. We reduce these potential problems by lagging the explanatories describing population composition; we use the 2006 values in the estimations of prices in 2009-2012.

When describing neighbourhoods, our basic unity is the census tract. I. e. we follow Denton and Massey (1991) when they *assume that they* [census tracts] *represent a reasonable approximation of a neighbourhood*. We use the characteristics of population composition in the tracts, as a core in our empirical analyses. One word of caution is important; the census tract is **a** reasonable approximation of a neighbourhood and clusters-of-tracts a reasonable

[4] Admittedly after some experimentation with different specification, which yielded inconclusive results. We did, for example, try out a specification where the effects of the concentration measure were allowed to vary with the level of the AA-variable.

definition of an extended neighbourhood; however, it is most definitely not necessarily **the** single most reasonable approximations.

In the empirical analysis, we are also utilising information at a clusters-of-tracts level. For some dimensions, the immediate neighbourhood matters most and for others an extended neighbourhood are more important. For example, schools cover more than one census tract and population composition at schools are important for house prices, Fiva and Kirkebøen (2011). The arguments for diversity are local in nature, Ottaviano and Peri (2006), but maybe not as restricted as a census tract. A final argument for using also clusters-of-tracts is that it enables us to improve considerably on our specifications. Within the aRE-model the use of two nested spatial levels enable us to distinguish between the effects of between and within variation of the population measures of interest. In the FE model, the census-of-tracts dummies control for unobserved attributes shared at the cluster-of-tract level.

4. Empirical background and data

We define Oslo and nine of its surrounding municipalities as the Oslo Metropolitan area. With a total number of 950,500 inhabitants, the Oslo Metropolitan area covers about 20 percent of the Norwegian population. Out of the total, 6.9 percent of the population in Oslo had, in 1990, an Asian or African background. In 2011, this share had grown to 17.2 percent. For more on this transformation, see Friedrichs et al. (2014). Real home prices in the Oslo region has been raising steadily for the last 20 years, as in most other parts of Norway. From 2000 to 2012, real prices of housing grew by 76.4 percent, an average annual real growth rate of about 4.8 percent.

Price data is taken from the net-based Housing for sale-portal, Finn.no, these data are updated with actual transaction prices when a unit is sold. We use data on transaction over the period

2009-2012. Note that the price variable consists of sales price[5]. Time trends in the data is captured by year-dummies – we regard the time span as short enough to be comfortable with the embedded hypothesis of time-invariant coefficients in the price equations. Finn.no covers more than 90 percent of the sales in the area. Information on population composition is taken from population registers, income distribution in neighbourhoods are taken from tax registers while information on the housing stock is taken from the 2001-census. We regard data from all these sources to be of very good quality.

Distances are measured by shortest travelling time by car in minutes between postal delivery zones. We account for speed limits. Employment data in each postal delivery zone used in the calculation of employment accessibility measure (4) is based on data from Statistics Norway.

[5] A large share of the Oslo housing stock consists of co-ops, where each housing units is responsible for its own part of a mutual debt. This mutual debt is added to the sales prices.

Table 1 – Descriptive statistics of the sample

	Median	P10	P25	P75	P90
Home prices Mill NOKs including cooperative debt.	2.482	1.537	1.863	3.600	5.250
Size in square meters	70	39	53	102	155
Year of construction	1966	1904	1940	1990	2007
Distance from CBD West (minutes)	9.6	3.4	4.6	14.7	21.6
Distance from CBD South (minutes)	14.5	10.5	12.3	19.3	23.5
Distance from CBD North-east (min)	10.7	5.5	6.8	17.3	22.8
Neighbourhood variables:					
Share Asians and Africans (AA)	0.109	0.041	0.056	0.187	0.321
Largest share single group in AA	0.306	0.205	0.246	0.428	0.579
Income	290	232	251	330	378
SD Income	200	133	161	378	641
Public housing shares	0.029	0.009	0.016	0.053	0.126
Single-family housing shares	0.059	0.006	0.013	0.268	0.386
Small units shares	0.291	0.134	0.182	0.474	0.608
Private rented units shares	0.197	0.108	0.143	0.326	0.420

5. Estimations and results

The two empirical models (the augmented random effects and the fixed effects model) described above has been estimated in order to shed light on the interdependencies between home prices and composition of the population. We present only the parameters of prime interest. The full estimations of the fixed effects model is placed in the Appendix. Other estimations results are available on request. They show e.g. that the coefficients of the not reported control variables are similar of both sign and magnitude in the two estimated models. One might note the combination of a relatively good fit (as measured by the R^2- measures) and by the fact that the signs of the coefficients of the variables that primarily act as controls are in line with a priori expectations and prior research, enhances our trust in the estimated models.

Table 2 – Regression results

	Augmented Random Effects		Fixed effects	
	Coefficients	SE	Coefficients	SE
AA-share	-0.211	0.060**	-0.207	0.011**
Mn-AA-share	-0.115	0.088		
Largest minority (LM)	0.039	0.018*	0.039	0.005**
Mn-LM	0.018	0.066		
LogIncome	0.130	0.040**	0.129	0.009**
Mn-LogIncome	0.254	0.120*		
SD Income	0.00018	0.00004**	0.00017	0.00002**
Mn- SD Income	0.00014	0.0001		
SD trunc dummy	-0.0021	0.013	-0.0016	0.003
R sq within	0.774		0.774	
R sq between	0.963		0.678	
R sq overall	0.0838		0.740	
Rho	0.084		0.483	
N=	98,568		98,568	

*Rho is the share of the residual variance that is due to the cluster-of-tracts level random intercept. Significance levels of 0.01 and 0.05 are marked with ** and *. SE denotes clustered standard errors.

First, consider the random effects model. The coefficients starting with the mn- prefix is the coefficients of the cluster-of tracts mean of the variables indicated after the prefix. As argued above they capture the difference between the within and between (cluster-of-tracts) effects of variations in the covariates. For the share of Africans and Asians in the neighbourhood, the size of the largest minority and for the standard deviation of the income there is no significant difference between the within and between effect. As the estimated coefficients of the within effects can be interpreted as instrumental variable estimates we have confidence in them, Rabe-Hesketh and Skrondal (2008). There is, however, a significant difference between the within and the between income effect.

Next we consider the coefficients of the within variation of the explanatories. All of them are significantly different from zero- and they are similar to the within estimates that the fixed effects model yields. This adds to our confidence in the estimated coefficients. Because of the similarities between the FE- and aRE estimates we will stick to the somewhat easier interpretable FE-estimates when we interpret and discuss our empirical findings.

The coefficient of the share of inhabitants with an African or Asian country background in the neighbourhood is -0.207. Roughly speaking this means that an increase from zero to one (i.e. all) decrease house prices by approximately 20 percent. This technical interpretation is however extrapolating far beyond the variation in our data and is, consequently not very interesting. The lower quartile in the distribution of the AA-share is 5.6, while the upper quartile is 18.7 percent. Our estimated coefficient predicts that prices decrease by 2,7 percent when the AA-share goes from the lower to the upper quartile. A movement from the 1st to the 9th decile (i.e. going from an AA-share of 4.1 to 32.1 percent) reduce prices by 5.6 percent.

Contrary to our prior belief, and the results of Ottaviano and Peri (2006) and Bakens et al. (2013), we do not find that a single dominant group within the population from Asia and Africa leads to lower prices than a diverse population from this regions does. Quite on the contrary house prices seem to be significantly increasing in the share that the single largest minority has in the AA-population of the neighbourhood. The estimated effect in our sample is, however, not large in magnitude. Using our estimates to predict prices at the 1st and 9th decile of the relative size of the largest minority increases prices, but only by 1.5 percent.

As we measure the neighbourhood (median) income by its natural logarithm in our empirical model, and our dependent variable is the log of the price of a housing unit, the coefficient can be directly interpreted as an elasticity. Consequently, a neighbourhood A with a one percent higher median male income than its neighbouring neighbourhood B, is expected to have a 0.13 percent higher home price level than neighbourhood B has.

The coefficient for income spread, as measured by the standard deviation of the income distribution, is significantly larger than zero. This is consistent with our hypothesis that most people (included home buyers) appreciate having different types of people around her – at least when it comes to income. In order to illustrate the magnitude of this effect we do again compare predictions at the upper and the lower quartile of the explanatory variable. Increasing the income diversity from a distribution with a SD of 161 (thousands of) NOKs up to 378 (i.e. a movement from the lower to the upper quartile) increase house prices by 3.8 percent. The model also include a coefficient of an indicator for whether the SD has been truncated or not (SD trunc dummy). Hence truncation does not seem to cause any harm to our estimations.

Our estimation results are (with one exception) consistent with the theory-based expectations. The magnitude of the impact from population composition into home prices is significant different from zero and substantially non-neglible, but we would add, they are by no means dramatic. The paper continues by first testing for any possible non-linearities in the responses towards variations in the share of inhabitants with an Asian or African country background in the neighbourhood, and then it proceeds by some sensitivity tests. Both the sensitivity tests and the test for non-linearities will done within in the FE-modelling frame.

6. Discontinuous price responses: Tipping?

In a much-cited article Card et al. (2008) demonstrate that majority population shares reacts in a highly non-linear fashion to changes in minority population. As the non-majority share passes through a *tipping point*, the majority share falls at an accelerating speed. The majority share falls rapidly because the housing demand of a majority of the majority population is directed elsewhere. It is natural to ask oneself whether prices of owner-occupied housing consequently declines at an accelerating speed as a major part of majority demand is withdrawn.

Card et al. (2008) argue that even if there exists a tipping point for the majority share, there need not be a tipping point in prices or rents. Their own empirical findings indicate that there is no tipping point in home prices (nor in rents) as the minority share passes the tipping point for majority shares. Still, we will test for break point in home prices.

Is there are any non-linearities in the relationship between the (log of) home prices and the share of minorities? To put it differently, are there any break points in the price response of the minority share? In order to test for this we follow a two-step procedure proposed by Card et al. (2008). First, calculate a series of price regressions containing a dummy for minority share above a threshold m', where m' is varied in short steps. If there exist a break point (or a tipping point) m*, the R^2 of these regressions will monotonically approach a maximum as one moves toward the break point, both from above and below. m* will coincide with a break point, if there exist a tipping point. Hence, m* is a candidate tipping point, but not necessarily a tipping point. Furthermore, if m* is well defined and is demonstrated not to be a tipping point, there does not exist any tipping points. Using this procedure, we find a candidate tipping point at an Asian-African share of 21 percent.

In order to test whether m* really is a tipping point include a spline at 0.21 in the FE-model, this tests for a gradual increase in the trend of falling prices as the AA-share passes through the candidate tipping point.

Table 3 – Regression results, price equations with break points

	Fixed effects, no break point		**Fixed effects, one break point**	
	Coefficients	SE	Coefficients	SE
AA-share	-0.207	0.011**	-0.163	0.021**
AA (AA>0.21)			-0.039	0.015**
Largest minority (LM)	0.039	0.005**	0.043	0.006**
LogIncome	0.129	0.009**	0.130	0.010**
SD Income	0.00017	0.00002**	0.00018	0.00001**
SD trunc dummy	-0.0016	0.003	-0.0021	0.003
R sq within	0.774		0.0774	
R sq between	0.678		0.676	
R sq overall	0.740		0.741	
Rho	0.483		0.482	
N=	98,568		98,568	

*Rho is the share of the residual variance that is due to the cluster-of-tracts level random intercept. Significance levels of 0.01 and 0.05 are marked with ** and *. SE denotes clustered standard errors.

The spline formulation gives a slope of the AA-share above the candidate tipping point that is significantly lower than the slope below the candidate tipping point. Hence, our empirical data enable us to reject a hypotheses of no tipping point in the interdependency between (the log of) home prices and the share of inhabitants with an Asian or African country background in the neighbourhood. To illustrate the magnitude of the 'break-point effect' we observe that increasing the AA-share with 5 percentage points from a level below the break point decrease house prices by 0.8%, while a similar jump from a point above the break point decrease prices by 1.1%. To our comfort we also observe that the other coefficients of interest are very similar in the models with and without the spline.

Price responses above and below the candidate tipping point is hence, quite similar in magnitude, even though the spline is significantly different from zero. It might be that it is a bit of an exaggeration to use the term tipping point in this context. However, some of the US experience, in particular during the late sixties and seventies make it appropriate to test for this in a setting where the immigrant population has grown strongly in some of the neighbourhoods of Greater Oslo.

7. Some sensitivity analyses

In order to assess the robustness of the empirical model we have estimated it on a number of subsets of the total sample. First, we estimate models separately for flats in blocks-of-flats and for single-family houses.

Table 4 – Regression results, separately by house types

	Blocks of flats		**Single family houses**	
	Coefficients	SE	Coefficients	SE
AA-share	-0.226	0.012**	-0.187	0.081*
Largest minority (LM)	0.061	0.007**	0.002	0.018
LogIncome	0.100	0.010**	0.170	0.038**
SD Income	0.0001	0.00001**	0.0002	0.00005**
SD trunc dummy	0.016	0.004**	-0.126	0.012
R sq within	0.740		0.366	
R sq between	0.447		0,675	
R sq overall	0.671		0.526	
Rho	0.580		0.405	
N=	77,014		9,866	

*Rho is the share of the residual variance that is due to the cluster-of-tracts level random intercept. Significance levels of 0.01 and 0.05 are marked with ** and *. SE denotes clustered standard errors.

Even though some small differences can be seen, the main impression is that the similarities between coefficients in the two market segments are remarkable – and they bear a close resemblance to the estimates of the FE-model estimated on the full sample. One can interpret Table 4 as a demonstration that the differences of the mechanisms studied in this paper is quite similar across market segments defined by house type. We then investigate whether this also holds for geographically divided markets. I.e. are the coefficients similar in the municipality Oslo to those in the wider Oslo housing market? What about the surrounding municipalities?

Table 5 – Regression results, separately for Oslo municipality and the surrounding municipalities

	Oslo municipality		**Only surrounding municipalities**	
	Coefficients	SE	Coefficients	SE
AA-share	-0.202	0.013**	-0.178	0.029**
Largest minority (LM)	0.063	0.008**	0.020	0.008*
LogIncome	0.158	0.012**	0.154	0.015**
SD Income	0.0001	0.0001**	0.0002	0.00002**
SD trunc dummy	0.014	0.004*	-0.036	0.006**
R sq within	0.788		0.775	
R sq between	0.768		0.384	
R sq overall	0.761		0.627	
Rho	0.463		0.610	
N=	72,193		26,375	

*Rho is the share of the residual variance that is due to the cluster-of-tracts level random intercept. Significance levels of 0.01 and 0.05 are marked with ** and *. SE denotes clustered standard errors.

Also for the estimations based on the transactions in the Central Oslo and for the periphery of the Wider Oslo housing market we find quite similar patterns. One may though note that the dummy identifying census tracts where the standard deviation of the income distribution has been truncated turns up with a significant coefficient in both the capital in the surrounding municipalities. However, we would say that these sensitivity exercises taken together do raise our confidence in the main estimates of the paper.

8. Concluding remarks

This paper tests how home prices in the Oslo region vary with central characteristics of neighbourhood population composition. Attractivity, bids and hence home prices is expected to be positively affected by neighbourhood affluence (measured by the median income of males 35-60 years old) and negatively affected by the share of Africans and Asians. Furthermore, we test for the correspondence between diversity in both income and country background. These tests are done within a hedonic price frame. We also find that income diversity feeds positively into house prices. Our tests do, however, not support the hypothesis that a single group dominating the AA-population give a lower price than a diverse AA-population does, quite on the contrary.

The results discussed above are all significantly different from zero in a statistical sense. We would still claim that the estimated effects are by no means of any dramatic magnitude. The same applies for our tests for tipping as the AA-share passes through a threshold. Even though price responses above a candidate point is significantly stronger than below, the difference is of a minor magnitude.

The paper reveals structures in home prices and sorting in the Oslo housing market. Next steps to be taken in order to gain further understanding of these processes is to facilitate studies that more explicitly addresses the questions of change and of causal links between population change and house price changes. Furthermore, there is a need to obtain data that links information on transacted properties and individual characteristics of the buyers. This would enable empirical analyses of group differences of the preferences of neighbourhood population composition.

Literature

Alesina A and La Ferrara E. (2005) Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43: 762-800.

Bailey MJ. (1959) Note on the economics of residential zoning and urban renewal. *Land Economics*: 288-292.
Bailey MJ. (1966) Effects of Race and of Other Demographic Factors on The Values of Single-Family Homes. *Land Economics* 42.
Bakens J, Mulder P and Nijkamp P. (2013) Economic impacts of cultural diversity in the Netherlands: Productivity, utility and sorting. *Journal of Regional Science* 53: 8-36.
Baumont C. (2009) Spatial effects of urban public policies on housing values. *Papers in Regional Science* 88: 301-326.
Bayer P and McMillan R. (2008) Distinguishing racial preferences in the housing market: Theory and evidence. *Hedonic methods in housing markets.* Springer, 225-244.
Blom S and Henriksen K. (2008) Levekår blant innvandrere i Norge 2005/2006 (Living Conditions among immigrants 2005/2006. *SSB Report.*
Card D, Mas A and Rothstein J. (2008) Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics* 123: 177-218.
Charles CZ. (2003) The dynamics of racial residential segregation. *Annual Review of Sociology*: 167-207.
Cheshire P and Sheppard S. (1998) Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics* 60: 357-382.
Chin T and Chau K. (2003) A critical review of literature on the hedonic price model. *International Journal for Housing Science and Its Applications* 27: 145-165.
Cutler DM, Glaeser EL and Vigdor JL. (1999) The Rise and Decline of the American Ghetto. *Journal of political economy* 107: 455-506.
Denton NA and Massey DS. (1991) Patterns of neighborhood transition in a multiethnic world: US metropolitan areas, 1970–1980. *Demography* 28: 41-63.
Dubin RA and Goodman AC. (1982) Valuation of education and crime neighborhood characteristics through hedonic housing prices. *Population and environment* 5: 166-181.
Farley R, Schuman H, Bianchi S, et al. (1978) "Chocolate city, vanilla suburbs:" Will the trend toward racially separate communities continue? *Social Science Research* 7: 319-344.
Fiva JH and Kirkebøen LJ. (2011) Information Shocks and the Dynamics of the Housing Market*. *The Scandinavian Journal of Economics* 113: 525-552.
Friedrichs J, Magnusson Turner L and Nordvik V. (2014) Migrant Diversity and segregation in Oslo 1990-2011. *ENHR.* Edinburg.
Galster G. (2003) Neighbourhood dynamics and housing markets. *Housing economics and public policy*: 153-171.
Galster G. (2012) *Driving Detroit: The quest for respect in the Motor City*: University of Pennsylvania Press.
Herath S and Maier G. (2010) The hedonic price method in real estate and housing market research. A review of the literature.
Hoff K and Sen A. (2005) Homeownership, community interactions, and segregation. *American Economic Review*: 1167-1189.
Li Q. (2014) Ethnic diversity and neighborhood house prices. *Regional Science and Urban Economics* 48: 21-38.
Moye R. (2014) Neighborhood racial–ethnic change and home value appreciation: evidence from Philadelphia. *Urban Geography* 35: 236-263.
Nathan M. (2012) After Florida: Towards an economics of diversity. *European Urban and Regional Studies*: 0969776412463371.
Osland L and Thorsen I. (2008) Effects on housing prices of urban attraction and labor-market accessibility. *Environment and planning. A* 40: 2490.
Ottaviano GI and Peri G. (2006) The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography* 6: 9-44.
Rabe-Hesketh S and Skrondal A. (2008) *Multilevel and longitudinal modeling using Stata*: STATA press.

Rosen S. (1974) Hedonic prices and implicit markets: product differentiation in pure competition. *The journal of political economy*: 34-55.
Schelling TC. (1971) Dynamic models of segregation†. *Journal of Mathematical Sociology* 1: 143-186.
Sethi R and Somanathan R. (2004) Inequality and segregation. *Journal of political economy* 112: 1296-1321.
Visser P, Van Dam F and Hooimeijer P. (2008) Residential environment and spatial variation in house prices in the Netherlands. *Tijdschrift voor economische en sociale geografie* 99: 348-360.
Wigren R. (1987) House prices in Sweden: the significance of attributes. *Scandinavian Housing and Planning Research* 4: 243-261.
Yinger J. (1976) Racial prejudice and racial residential segregation in an urban model. *Journal of Urban Economics* 3: 383-396.
Yinger J. (2014) Hedonic Estimates of Neighborhood Ethnic Preferences. *Public Finance Review*: 1091142113520562.
Zabel JE. (2008) Using hedonic models to measure racial discrimination and prejudice in the US housing market. *Hedonic methods in housing markets.* Springer, 177-201.
Zhang J. (2011) Tipping and Residential Segregation: A Unified Schelling Model. *Journal of Regional Science* 51: 167-193.

Appendix

Fixed effects regression of log home prices, fixed effects at the cluster of tracts level

	Coefficient	SE
Log size single family houses	0.588	0.006**
Log size flats	0.644	0.006**
Log size Other house types	0.772	0.002**
Construction year -1940, dummy	Ref	
Construction year 1940-50, dummy	0.003	0.005
Construction year 1951-70, dummy	-0.032	0.002**
Construction year 1971-80, dummy	-0.028	0.003**
Construction year 1981-90, dummy	0.001	0.003
Construction year 1991-2000, dummy	0.049	0.003**
Construction year 2001-2006, dummy	0.127	0.003**
Construction year 2007-, dummy	0.130	0.003**
Owner occupied, dummy	Ref	
Co-op, dummy	-0.002	0.002
Share in housing company, dummy	-0.009	0.003**
Flat ground floor, dummy	Ref	
Flat 1st or second floor, dummy	0.027	0.002**
Flat 3rd or 4th floor, dummy	0.052	0.002**
Flat above 4th, dummy	0.083	0.003**
Sold 2009, dummy	Ref	
Sold 2010, dummy	0.083	0.002**
Sold 2011, dummy	0.182	0.002**
Sold 2012, dummy	0.251	0.002**
Travel time in minutes from CBD, to the west	-0.002	0.0007**
Travel time in minutes from CBD, to the north-east	-0.007	0.001**

Travel time in minutes from CBD, to the south	-0.0005	0.0004
Log accessibility	-0.021	0.006**
Block of flats, dummy	Ref	
Single family house, dummy	1.00	0.032**
Row- or terraced house, dummy	0.652	0.031**
Two dwelling houses, dummy	0.662	0.033**
Share public rental housing, neighbourhood	0.112	0.012**
Share single family houses, neighbourhood	-0.041	0.007**
Share blocks of flats, neighbourhood	-0.024	0.004**
Share small units	-0.012	0.009
Share private rental units, neighbourhood	0.073	0.009**
Share AA, neighbourhood	0.211	0.011**
Share largest minority, neighbourhood	0.039	0.005**
Log median male income, neighbourhood	0.130	0.009**
SD median male income, neighbourhood	0.0002	0.00001**
Indicator truncated SD, neighbourhood dummy	-0.002	0.003
Intercept	10.814	0.096
R^2- within	0.774	
R^2-between	0.674	
R^2-overall	0.740	
Rho	0.483	
N=	98,568	

Significance levels of 0.01 and 0.05 is marked with ** and *. SE denotes clustered standard errors.

Note that neighbourhood in the table above refer to census tracts.