ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis Economics of Sustainability

Segregation Preferences and Socio-Economic Status:

Evidence from Rotterdam Social Housing

Armin Hoendervangers
428365

Supervisor: Dr. Yao Chen
Second assessor: Dr. Anna Baiardi
Date: October 26, 2023

Abstract

This thesis investigates how socio-economic status of current residents shape prospective residents' preferences for segregation. Combining a novel dataset on Rotterdam social housing with administrative data, I use a modified hedonic pricing model to estimate (i) prospective residents' preferences for less segregation and (ii) the effect of *current* residents' socio-economic status on these preferences. I find that, in general, prospective residents have a postive preference for less segregation. Of the socio-economic indicators considered, elderly age, household size, household income, and education level appear to have a significant and negative effect on the desirability of less segregation.

Contents

1 Introduction

As the governing body of a diverse and multicultural city, one of the main goals of the municipality of Rotterdam is to combat discrimination in – amongst other things – the housing market (Gemeente Rotterdam, 2020; Simons et al., 2022). In this same agreement, the municipality claim they aim to continue using selective assignment of social housing based on the Dutch *Wet Bijzondere Maatregelen Grootstedelijke Problematiek*, which allows for preferential treatment in the social housing market based on socio-economic status, which potentially induces socio-economic segregation. Research has shown that socio-economic status is one of the factors shaping preferences for ethnic segregation (Clark, 2009; Iceland & Wilkes, 2006); preferential treatment based on socio-economic status might thus be counterproductive with respect to tackling ethnic segregation.

Doucet and Koenders (2018) describe the experience of the residents of neighbourhoods targeted by such policies aimed at increasing "social mix". Although attitudes are mixed, the authors note that many of their interviewees appear to prefer diverse neighbourhoods. As one interviewee puts it:

> [...] I see that whites are coming back to the Afrikaanderwijk. I think that's good, I really do. First it was a ghetto. [...] all these [foreigners] living together. Where were all the white people? They had migrated to the suburbs and now they are coming back. I think that a diverse society is really good.

In other words, it seems that the current residents of Rotterdam neighbourhoods prefer less segregation.

Social housing is one of the key factors – and thus policy instruments – shaping urban segregation (Verdugo & Toma, 2018). To successfully make use of this instrument, policymakers should tailor policy to prospective residents. Although preferences for segregation are shaped by socio-economic status, the effect the socio-economic status of *current* residents has on these preferences remains unclear. This thesis therefore addresses the following research question:

What is the effect of a neighbourhood's socio-economic status on prospective residents' preferences for less segregation in Rotterdam social housing?

This thesis' contribution to the existing literature is twofold. First, I exploit novel data to investigate the relationship between socio-economic status and segregation preferences in the distinct setting of social housing, rather than the unregulated (non-)rental market. Second, I examine the effect the socio-economic status of current residents has on prospective residents' preferences for segregation, a venue to the best of my knowledge yet unexplored by current literature.

I make use of a modified two-step hedonic pricing model to estimate the marginal value of less segregation and how this is affected by socio-economic status of current neighbourhood residents. Using two measures as a proxy for home value, I examine both preferences and willingness to pay for less segregation. Data on social housing adverts is combined with publicly available aggregated spatial data from Statistics Netherlands. To increase precision of available variables, I attempt to disaggregate data using three levels of spatial units with non-overlapping borders. Finally, I check whether results are robust to the weights used for disaggregation and the definition of segregation used.

I find that several socio-economic aspects of the neighbourhood's current residence affect the perception of less segregation for that neighbourhood: the share of elderly residents, the average household size, household income, and the share of residents with a lower education level. The share of children in a neighbourhood and the share of residents that experience difficulties with the Dutch language also appear to affect preferences for less segregation, but do not show a significant effect on willingness to pay.

The rest of this thesis proceeds as follows. Section 2 provides an overview of related literature. Section 3 describes the Rotterdam social housing system, Section 4 the data used, and Section 5 the empirical framework. Sections 6 and 7 present the main results and robustness checks, respectively. Finally, Section 8 concludes.

2 Literature Review

Existing literature on segregation preferences present contrasting conclusions. Zhang and Zheng (2015) use a migration choice model to show that segregation is an urban disamenity. They find that segregation reduces always reduces utility, although this varies based on socio-economic status and ethnicity. Ibraimovic and Hess (2018) and Ibraimovic and Masiero (2014), on the other hand, find a positive preference for segregation, originating from both a preference for living with co-nationals and a preference for a lower share of foreigners. Similar to Zhang and Zheng, however, they also find that the strength of these preferences depends on socio-economic status and nationality.

Most of current literature takes into account at least one of socio-economic status and ethnicity when examining segregation preferences. With respect to ethnicity, in general, results suggest prospective residents prefer to live among those which share their ethnicity (Clark, 2009; Daniels, 1975). The strength of this preference is not equal across ethnicities, however, and also appears to depend on socio-economic status.

Literature examining how socio-economic status shapes segregation preferences also shows conflicting findings. Järv et al. (2021) use spatial activity and survey data to examine how socio-economic status affects ethnic segregation. Although results are not straightforward, they find that individuals in higher social classes tend to self-segregate. In contrast to this, based on a longitudinal analysis Hwang et al. (1985) find that socio-economic status is not a significant explanatory factor of segregation. Finally, another strand of literature shows how (preferences for) segregation decreases as socio-economic status increases (e.g. Clark, 2009; Iceland & Wilkes, 2006; Spivak & Monnat, 2013).

To determine or estimate preferences for segregation, the literature generally follows one of two approaches. The first strand makes uses of stated preferences, where respondents directly make their preferences known (e.g. Clark, 2009; Ibraimovic & Masiero, 2014). The second strand makes use of revealed preferences, through a home value-based approach (e.g. Daniels, 1975; Zhang & Zheng, 2015). This thesis follows this second strand of literature, as I estimate preferences for segregation through a hedonic pricing model using a proxy for home value. As follows from the research question, I focus on the influence of socio-economic status on preferences of prospective residents. As data on the ethnicity of prospective residents is not available, the influence of ethnicity is precluded from my analysis. Nevertheless, this thesis adds to the literature through novel data on Rotterdam social housing, providing an image of any differences in preferences within a group with relatively lower socio-economic status. Furthermore, in contrast with existing literature, I examine the role of socio-economic status of current residents rather than that of prospective residents, which to the best of my knowledge has not been examined before.

3 Institutional Background

Social housing in Rotterdam is allocated based on four different methods: registration length, lottery, first come first serve, and "Wish and Wait".

In the registration length method, The applicants are sorted based on the time since registering in the social housing system, so that the applicant that registered the longest ago is offered the home first. In the lottery method, the sorting order of applicants is randomised, so that all applicants have an equal chance of getting offered the home. In the first come first serve method, as implied by the name, the applicants are sorted based on when they respond to the advert, with the first respondent being offered the home first. Finally, in the "Wish and Wait" method, applicants apply for a *cluster* of homes. By applying for such a cluster, the applicants register for a cluster-specific waiting list. Once a home in the cluster is free, the first on the waiting list is offered the home.

In addition to these allocation methods, applicants in emergency situations can also be given "urgency" status. Those with urgency status get priority over non-urgency applicants in the registration length and lottery methods. Furthermore, should these applicants not be successful despite the urgency status, the housing associations will look for a home for these applicants through so-called mediation.

Finally, for some specific homes applicants can also be given priority through the so-called *Rotterdamwet*. The Rotterdamwet is a Dutch law that allows municipalities to give priority to individuals with certain socio-economic characteristics in the allocation of social housing, such as being employed or doing volunteer work.[1] Adverts that make use of the Rotterdamwet are all allocated based on applicant registration length, with priority being given to applicants that fulfill the stated Rotterdamwet "requirements" for that advert.

An applicant's registration length is determined by the time at which they registered in the WoonnetRijnmond application system. Note that this is different from applying for a home: registering enables applicants to apply on adverts for specific homes. The system has a yearly registration fee and registration length builds up as long as this yearly fee is paid. Once an applicant moves to a different independent home within the Netherlands, their registration length is set to zero – even if this home is not obtained through WoonnetRijnmond.

4 Data

For this thesis I collect novel data on all filled social housing adverts in the municipality of Rotterdam from the WoonnetRijnmond website. As of writing, observations span from 17th of June 2022 up to 15th of June 2023. For all observations, I obtain the home address, allocation method, the registration length of the successful applicant, and the total amount of responses.

I combine this data with geographical data from Statistics Netherlands on three different observation levels: 5 digit postal code, neighbourhood, and 500m^2 square areas. Although each of these areas may contain several homes, the boundaries of the units do not exactly overlap, so that there is more variation between homes compared to using just a single of these units.[2] I obtain demographic and socio-economic information for each of these areas, such as the average WOZ building value, the age distribution, average household size, and origin of the population.

Third, from the municipality of Rotterdam's *Wijkprofiel Rotterdam*, I obtain data on the amount and accessibility of amenities on the neighbourhood level.

Finally, from Kadaster (the Netherlands' Land Registy and Mapping Agency) I obtain the building year and surface area for each of the homes in the first dataset. I combine these datasets using the geographical coordinates associated with the addresses in the first dataset, and match each address with its respective area or observation using the Geographic Information System software QGIS.

4.1 Variable Disaggregation

The Statistics Netherlands geographical datasets contain mostly the same variables, but aggregated on a different observation level. In order to obtain as accurate as possible

[1] For more information on the Rotterdamwet, see https://www.woonnetrijnmond.nl/service-en-contact/regelgeving/wat-is-de-rotterdamwet-en-een-rotterdamwet-woning/ (in Dutch).

[2] See Appendix Figure A1 for an example of how the boundaries look in practice.

values for each variable, I use a weighted average to disaggregate the values, where each variable is weighted by the inverse of its population measure. This way, the more granular an observation level is, the more weight it has. As an example, when disaggregating the share of elderly in a population, the inverse of the total amount of population of the observation level is used. Each aggregated variable x is thus calculated in the following manner:

$$x_i = \frac{\sum_u n_{i,u}^{-1} x_{i,u}}{\sum_u n_{i,u}^{-1}} , \tag{1}$$

where i and u index observation and observation level respectively, and n denotes the population measure, i.e. the level's total population, amount of households, or amount of homes.

In addition to these weighted averages, I also calculate a simple average over observation levels to check for robustness of results:

$$x_i = \frac{1}{U} \sum_{u=1}^{U} x_{i,u} . \tag{2}$$

In the rest of this thesis, variables obtained using Equation (1) are suffixed with "(inv)", while variables obtained using Equation (2) are suffixed with "(avg)", unless indicated otherwise.

4.2 Measuring segregation

To quantify segregation, I follow Reardon and Firebaugh (2002) and conceptualize segregation as low diversity or as disproportionality. The notation used in this subsection is as follows: p denotes proportion in an area, π denotes total or city-wide proportion, subscript i indexes geographical areas, and subscript j indexes ethnic groups. Therefore $p_{i,j}$ denotes the share of area i's population of ethnic group j, and π_j denotes the share of the total (city-wide) population of ethnic group j.

Diversity When evaluating segregation as a lack of diversity, segregation can be measured using the information theory index H. As this is derived from the diversity conceptualisation, from here on I will denote this index as diversity index D. The diversity index for a given area i is computed as follows:

$$D_i = -\sum_{j=1}^{J} p_{i,j} \ln p_{i,j} , \tag{3}$$

where $p_{i,j}$ is the proportion of area i's population with ethnicity j. In case the proportion of a given ethnicity is zero, that ethnicity's contribution to the index ($p_{i,j} \ln p_{i,j}$) is defined as zero. The maximum value of the index depends on the amount of groups J considered, and is equal to $\ln J$ when the proportion of all J groups is equal. For ease of interpretation, I normalise the index by dividing it by its maximum possible value and multiplying it by one hundred, so that $D_i \in [0, 100]$. A value of zero then denotes full segregation, i.e., the entire population of area i is of a single ethnicity, while a value of one hundred denotes no segregation, where all ethnicity groups in area i have an equal proportion of the population.

Disproportionality The disproportionaly approach is based on the ratio between the local and total proportion of an ethnic group. The ratio r for ethnic group j in area i is defined as:

$$r_{i,j} = \frac{p_{i,j}}{\pi_j} \ . \tag{4}$$

Segregation in area i is then zero if and only if $\forall j \in k : r_{i,j} = 1$, i.e. the proportion of each ethnic group in area i is equal to the proportion of that same group in the total population. In other words, area i's population is fully representative (in terms of ethnicity) of the total population. If any $r_{i,j}$ is not equal to one, segregation is larger than zero. For ease of comparison with the diversity-based index, let representation be the opposite of segregation, so that zero segregation is the same as full representation (a value of one). A representation index can then be defined as a weighted average of the representation of each ethnic group in an area, where representation is measured by some function $f(r_{i,j})$. By construction the range of r is

$$\frac{0}{\pi_j} = 0 \leqslant r_{i,j} \leqslant \infty = \lim_{\pi_j \to 0} \frac{1}{\pi_j} \ . \tag{5}$$

The function $f(\cdot)$ should thus satisfy the following properties:

$$f(0) = 0 \ , \qquad \lim_{r_{i,j} \to \infty} f(r_{i,j}) = 0 \ , \qquad f(1) = 1 \ ,$$
$$f'(1) = 0 \ , \qquad f'(r_{i,j}) < 0 \ \forall r_{i,j} \in (1, \infty) \ , \qquad f'(r_{i,j}) > 0 \ \forall r_{i,j} \in (0, 1) \ .$$

In other words, the function reaches its maximum value of one when the ratio is one, the function is strictly decreasing as the ratio deviates from one, and the function is zero when the proportion of an ethnic group is zero. In this thesis I use the following function $f(\cdot)$ which satisfies all of these properties:

$$f(r_{i,j}) = r_{i,j} \exp(1 - r_{i,j}) \ . \tag{6}$$

The representation index is then the sum of these values, weighted by the population share of each ethnic group:

$$R_i = \sum_{j}^{k} \pi_j f(r_{i,j}) \ . \tag{7}$$

The smallest possible value of this index is obtained when an area consists solely of the ethnic group with the smallest total proportion. This index will therefore always be larger than zero, although its minimum value will approach zero as the smallest total proportion approaches zero. This fits conceptually: in each area, there will always be at least one group that is represented by the local population. If no group is represented the proportion of each group must be zero, which is impossible; every individual belongs to a group, so that not every proportion can be zero as long as the population is larger than zero.

As with the disproportionality index, I multiply the index by one hundred so that $R_i \in (0, 100]$.[3] Values closer to zero denote then denote more segregation, whereas a value of one hundred denotes no segregation, i.e. complete representation.

4.3 Home value

In this thesis I use two proxies to measure the value of a home: the registration length of the successful applicant and the total amount of responses for a housing advert.

The successful applicant's registration length can be considered analogous to the price of a home in the non-rental market. When buying a new home, prospective residents have a set of options (their choice set) available based on the mortgage available to them (their budget constraint). When multiple individuals are interested in a home that is for sale, the home is sold to the highest bidder. The choice set is therefore limited to homes where prospective residents would be the highest bidder. I assume prospective residents to be rational, so that the home they choose from their choice set is the home that is most valuable to them, i.e. that best fulfills their preferences. In the case of social housing, the budget constraint is not a mortgage but a prospective resident's registration length; as explained in Section 3, in the main allocation methods, the applicant with the longest registration time is offered the home. Based on this analogy, I view the succesful applicant's registration length as a price "paid" to obtain the home. Similar to a regular "price", should an applicant accept the offer for the home, they have to "give up" their current registration length: registration length is set to zero whenever an applicant moves to a different home within the Netherlands.

The other proxy, the amount of responses for an advert, is a more straightforward measure of a home's popularity. Simply put, the higher people value a given home, the more likely they are to apply for it. Thus, the higher the amount of total applicants, the higher the value of the home can be considered to be.

The interpretation of each proxy is slightly different. As registration length is "paid" for a home, a model using registration length analyses the willingness to pay for less

[3] Note that the lower bound will be larger than zero, as it depends on the groups considered in the analysis. Theoretically, as one considers more – and therefore smaller – different groups, the lower bound tends to zero.

segregation (in terms of registration length). A response on an advert, on the other hand, is not “paid”: a response can be retracted if an applicant wishes, and current responses have no effect on the amount of responses later. A model using responses therefore analyses applicant preferences rather than willingness to pay. The results of the two different kinds of models could therefore show whether preferences and willingness to pay for less segregation align.

4.4 Amenities and welfare indices

To quantify the availability of amenities, I construct indices for different types of amenities using principel component analysis. Each index is defined as the first principal component of the used variables, so that the index is a linear combination of the standardised variables. The loadings of the first principal component then correspond to the weight given to each variable. The variables for each amenity index are selected in such a way that a higher score on the index corresponds to that type of amenity being more readily available. As the variables are standardised before being combined into the indices, the average value for each index is zero.

I construct indices for the availability of stores, health services, hospitality services, leasure activites, schools, and public transport. In addition to these amenities, I also construct an index to measure the prevalence of welfare usage, as an indicator of socio-economic status. The weights assigned to the underlying variables as well as the distribution of the resulting indices are reported in Appendix B.

4.5 Summary statistics

All observations are allocated to one of four submarkets based on their geographical location as part of the emperical framework (see Section 5). As such, summary statistics are additionaly reported per submarket. Figure 1 depicts the geographical location of each submarket.

Table 1 reports summary statistics for each submarket and over all observations. The overall average registration length of a succesful applicant is approximately 1187 days, or 39 months.[4] Registration length is higher in the West submarket compared to the Centre, North, and South submarket, where registration length is closer to the overall average. The average amount of responses to a home advert is approximately 440, with a higher average in the Centre and South submarket of approximately 520. Both registration length and responses have a high standard deviation as values are rather spread out. In the centre and south of Rotterdam, the diversity index shows relatively high values, which fits the multicultural image of Rotterdam. In the North and West submarkets, however, diversity appears to be lower. Values for the representation index are higher for

[4] This includes homes not allocated based on registration length, and is therefore likely lower than figures reported in popular media discussing social housing.

Table 1: Summary Statistics by Submarket

	Centre	North	South	West	Total
Registration length (days)	1249.16	962.15	1144.03	1687.63	1187.40
	(1269.76)	(1162.58)	(1091.46)	(1490.58)	(1224.39)
Responses	519.66	345.60	524.57	223.36	440.18
	(665.90)	(618.42)	(623.32)	(247.17)	(609.46)
Diversity (inv)	76.68	60.45	72.88	50.24	67.75
	(6.40)	(10.89)	(10.65)	(18.75)	(14.39)
Representation (inv)	86.39	85.28	84.66	77.79	84.37
	(9.81)	(8.44)	(10.78)	(14.00)	(10.78)
Surface area	78.06	67.72	70.53	79.24	72.58
	(23.13)	(14.97)	(15.86)	(23.43)	(19.11)
Home age	71.03	48.88	60.70	46.70	58.28
	(36.56)	(14.50)	(26.61)	(17.47)	(27.41)
WOZ value (inv)	261.65	245.38	200.66	227.57	229.46
	(46.85)	(55.23)	(41.04)	(39.19)	(52.55)
Amenities					
Stores and shops	2.48	−1.44	−0.04	−1.44	0.00
	(1.30)	(1.29)	(1.94)	(0.57)	(2.13)
Health services	2.37	−0.42	−0.18	−2.97	0.00
	(0.83)	(1.11)	(1.67)	(1.07)	(2.03)
Hospitality	1.46	−1.07	0.02	−0.49	0.00
	(0.52)	(1.80)	(1.23)	(1.24)	(1.58)
Leasure activities	1.79	−1.48	0.96	−3.29	0.00
	(0.64)	(1.21)	(0.90)	(0.70)	(1.96)
Schools	1.48	−0.54	0.32	−2.63	0.00
	(0.84)	(1.89)	(1.51)	(2.55)	(2.06)
Public transportation	1.21	−0.07	0.62	−4.10	0.00
	(0.35)	(0.60)	(0.54)	(1.22)	(1.72)
Neighbourhood safety	108.22	127.16	103.53	132.69	114.29
	(11.35)	(9.90)	(12.11)	(6.88)	(15.99)
Socio-economic status					
Share >65 (%, inv)	14.22	30.04	17.11	27.06	21.03
	(7.76)	(12.34)	(11.66)	(11.71)	(12.85)
Share 0–14 (%, inv)	14.69	13.76	18.72	14.08	15.95
	(4.45)	(5.33)	(5.42)	(4.72)	(5.57)
HH size (inv)	1.90	1.82	2.04	1.97	1.94
	(0.27)	(0.31)	(0.23)	(0.26)	(0.28)
HH income (×1000€, inv)	24.48	25.97	23.00	27.12	24.61
	(3.74)	(4.25)	(2.71)	(3.29)	(3.78)
Welfare usage	−0.41	−0.51	1.06	−1.53	0.00
	(2.14)	(1.37)	(1.43)	(0.98)	(1.81)
Share language barrier (%)	19.21	9.65	16.76	7.68	14.36
	(4.87)	(2.21)	(5.82)	(2.66)	(6.29)
Share no start qual. (%)	31.27	27.80	38.93	34.04	33.70
	(10.07)	(6.19)	(4.14)	(1.45)	(7.81)
Observations	923	1050	1582	496	4051

Notes: Standard deviation in parentheses.

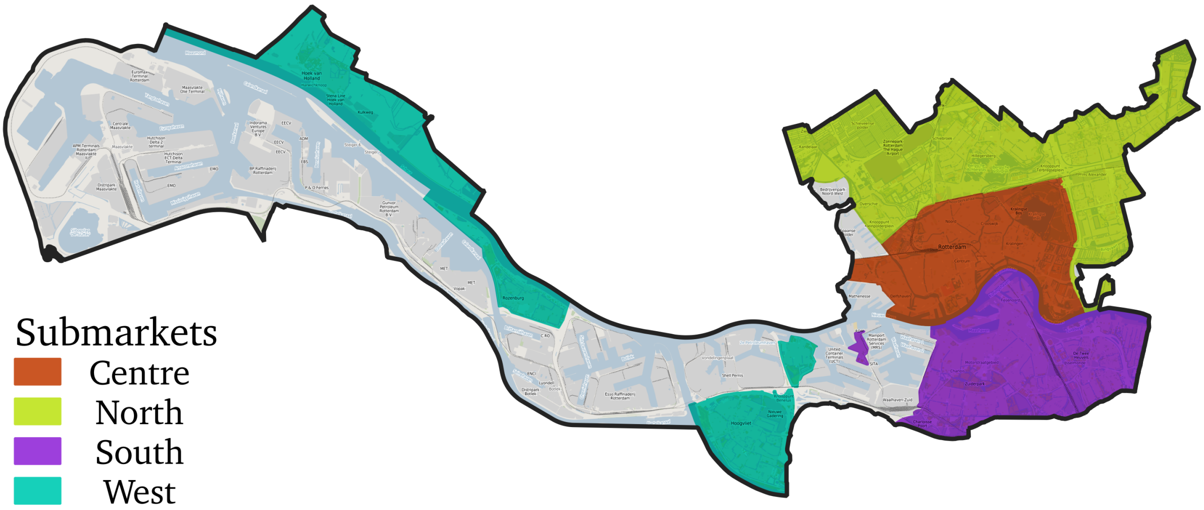

Figure 1: Submarket Allocation
Notes: Map depicting geographical location of the four used submarkets. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table 2: Allocation Method by Submarket

	Centre	North	South	West	Total
Registration length	0.32 (293)	0.54 (562)	0.31 (484)	0.79 (391)	0.43 (1730)
Lottery	0.22 (199)	0.09 (91)	0.14 (220)	0.07 (34)	0.13 (544)
DirectKans	0.12 (115)	0.16 (166)	0.15 (235)	0.02 (11)	0.13 (527)
Wens&Wacht	0.00 (1)	0.03 (31)	0.15 (234)	0.00 (0)	0.07 (266)
Rotterdamwet	0.06 (57)	0.00 (0)	0.05 (82)	0.00 (1)	0.03 (140)
Urgency	0.28 (258)	0.19 (200)	0.21 (327)	0.12 (59)	0.21 (844)
Observations	923	1050	1582	496	4051

Notes: observation count in parentheses.

each submarket, with the lowest average in the West submarket. Interestingly, the North submarket scores similar to the Centre and South submarket in terms of representation, even though its diversity score is lower. Most homes are approximately equal in size, with an average surface area of approximately 73 square metres.

Table 2 reports the share of each allocation method used for each submarket. For each submarket, registration length is the most used allocation method, followed by urgency. The prevalence of the lottery, first come first serve, “wish and wait”, and Rotterdam act methods differs per submarket.

5 Empirical Framework

To estimate the effect of socio-economic status on segregation preferences, I base my methodology on Rosen’s (1974) hedonic pricing method. I use registration length and advert responses as a proxy for home prices, as described in Section 4. The main model described below is thus estimated twice: once using registration length as the price proxy, and once using responses as the price proxy.

5.1 Hedonic regression

First, I estimate a model where the hedonic price, or value, of home i depends on the quantity and quality of its characteristics and amenities:

$$\ln y_i = \alpha + \beta D_i + \sum_j \gamma_j x_{i,j} + a_i + \varepsilon_i \ , \qquad (8)$$

where the dependent variable y_i is the value of home i, D_i is the diversity index of the home's geographical area, $x_{i,j}$ is the jth characteristic or amenity of the home, a_i is a dummy variable for allocation method, and ε_i is an idiosyncratic error term. The coefficients then show how much each of the aspects add – or detract – from the value of a home. The model is estimated in log-linear form so the effect of each characteristic on the value of a home is nonlinear. This nonlinearity allows for variation in the marginal value of a characteristic, allowing for further analysis in the next steps of the empirical framework.

Not all allocation methods are controlled for by a_i as not every allocation method can be expected to affect the amount of responses or the registration length of the succesful applicant. For the responses models, I control for the lottery, first come first serve, wish and wait, and Rotterdam act methods. These methods are observed on the home advert and could influence the likelihood of an applicant responding, as an applicant might perceive their odds of success to be higher for certain methods. For example, applicants that do not fulfill the requirements posed in a Rotterdam act advert might decide not to respond to that advert at all; they know they will not receive priority on that advert's waiting list, so using their limited amount of open responses on another advert might be a more fruitful endeavour. I do not control for the urgency method as that is a status that an applicant might have. Other applicants do not – and cannot – know beforehand whether a home will go to an applicant with urgency status, so this does not affect an applicant's *ex ante* perceived odds of success.

For the registration length models, I control for the wish and wait, Rotterdam act, and urgency methods as these have a direct effect on the amount of registration length of the successful applicant. The priority granted to applciants by the Rotterdam act or urgency status lowers the amount of registration length needed for their respective applicants to get to the top of the waiting list for a home: these applicants get priority over applicants that do not fulfill the respective requirements, even if the applicants without priority have a higher amount of registration length. On the other hand, the wish and wait method is used for homes that are not directly available. Applicants that apply can thus expect to wait for a while longer before that home becomes available – adding to their registration length. Adverts allocated using the lottery and first come first server method are excluded from these models, as these homes are not allocated based on registration length: these observations therefore do not add any useful information to the model. Finally, for both sets of models the "regular" registration length allocation method is used as the reference category.

5.2 Computing marginal value added

After estimating the hedonic regression model, I define the revealed preference for less segregation as the marginal value added by the entropy index. The marginal value added due to less segregation is computed as follows:

$$mv_i = \frac{\partial y_i}{\partial D_i} = \hat{\beta} \exp(\hat{\alpha} + \hat{\beta} D_i + \sum_j \hat{\gamma}_j x_{i,j} + \hat{a}_i) \; , \tag{9}$$

where mv_i is the marginal value of less segregation in terms of registration length or additional responses.

5.3 Preference model

Generally, the two-step hedonic pricing method is used to estimate a structural demand parameter (see, e.g., Mei et al., 2017; Poudyal et al., 2009). There are two concerns with the second step of Rosen's method.

First, Bartik (1987) and Palmquist (1984) argue that when a non-linear hedonic regression is used, a household's choice for the quantity of a characteristic implies a simultaneous choice of the marginal price of that characteristic, as the marginal price is endogenous: it depends on the quantity of the characteristic. Furthermore, unobserved preferences influence a household's choice in both the quantity and marginal price of a characteristic. The traditional approach to address is this issue is the use of instruments for the estimated marginal price, specifically variables that would shift the demand function, i.e. variables that proxy for unobserved preferences.

As the goal of this thesis is not to estimate a demand parameter, but rather to investigate preferences for less segregation, the interest lies in precisely the first stage of the IV-regression, where the marginal price is regressed on variables that proxy for unobserved preferences. By leaving the characteristic's quantity – in this case the amount of segregation – out of the second model, the issue of endogeneity is circumvented.

The second concern, raised by Brown and Rosen (1982), is that the marginal prices derived from the observed quantity of a characteristic do not provide any additional information in themselves. A second regression on the marginal prices would therefore simply "reproduce" the information provided by the first regression. A common solution to this issue, proposed by both Brown and Rosen (1982) and Palmquist (1984), is to segment the market in spatially distinct submarkets and estimate the first regression for each of these submarkets separately. The second regression is then estimated on the market as a whole, pooling the observations from each submarket. The underlying assumption is then that structural demand is the same across all submarkets, while unobserved submarket-specific factors are reflected in the coefficient estimates. As the marginal prices depend on these "local" estimates, additional information is obtained by pooling all observations. To address this second concern, I follow the outlined procedure and estimate the first regression separately for each of Rotterdam's districts. I then estimate a second city-wide model to investigate the influence of socio-economic factors on the marginal value added due to less segregation:

$$\operatorname{asinh} mv_i = \delta + \sum_k \eta_k s_{i,k} + w_i + a_i + \omega_i \ , \tag{10}$$

where $s_{i,k}$ is the kth socio-economic factor measured in the area of home i, w_i are district fixed effects accounting for unobserved neighbourhood characteristics, a_i are dummy variables accounting for the allocation method as in Equation (8), and ω_i is an idiosyncratic error term.

For the dependent variable, I use the inverse hyperbolic sine (IHS) of marginal value added as this approximates a log-linear model, but is also defined for negative values.[5] A caveat of the IHS transformation is that the obtained coefficients cannot be interpreted as precentage changes as easily as when using a log-transformation (Bellemare & Wichman, 2020). The percentage change is approximated by the calculation used for log-linear models for large values of the dependent variable, but diverges as the values approach zero. For small values, the magnitude of the percentage change should thus be viewed only as an approximate indication of the actual magnitude of the percentage change. In spite of this downside, I make use of the IHS transformation as it allows both for negative outcomes as well as modelling a multiplicative relationship between independent and dependent variables. Furthermore, I interpret the results using the log-linear approximation of a percentage change, noting that the approximation is biased downwards for small values (< 10) of the dependent variable (Bellemare & Wichman, 2020).

5.4 Robustness checks

To check whether results are sensitive to used definitions, I perform two robustness checks. First, I re-estimate the main models using a measure of disproportionality rather than diversity to quantify segregation. For this robustness check, I use the representation index (Equation (4)) rather than the diversity index. The use of a different operationalisation of segregation then shows whether results are dependent on the used measure of segregation.

[5] IHS is defined as $\operatorname{asinh} y = \ln(y + \sqrt{y^2+1})$. For large values of y, $y + \sqrt{y^2+1} \approx 2y$, so that $\operatorname{asinh} y \approx \ln(2y)$. For values of y close to zero, $\operatorname{asinh} y \approx y$.

Second, I re-estimate the main models using the simple average disaggregated variables rather than the inversely weighted averages. The results of this robustness check should then show whether the main models' results are sensitive to a different choice of weights for the aggregated variables. Robustness for disaggregation is also checked in Appendix C, using non-disaggegated variables.

6 Results

6.1 Hedonic model

Table 3 reports the estimated coefficients of the hedonic regression model (Equation (8)) using the amount of responses for an advert as the dependent variable, where each column corresponds to one of the previously defined submarkets.

For all submarkets, there appears to be a positive association between the amount of responses for a home advert and the ethnic diversity in that home's area, all statistically significant at the 0.1 percent level. A one unit increase in diversity corresponds to an increase in responses between approximately 1.85 (Northern submarket) and 2.91 (West submarket) percent,[6] ceteris paribus.

Table 4 reports the estimated coefficients of the hedonic regression model using the registration length of the succesful applicant as the dependent variable, where each column corresponds to one of the previously defined submarkets.

In contrast to the responses model, only the south and west submarket show a significant relationship between diversity and registration length, at the 1 percent level. For both these two and the centre submarket, the relationship is positive, while the relationship is insignificant and negative for the north submarket. A one unit increase in diversity corresponds to a change in registration length ranging from a 0.46 percent decrease (North submarket) to a 1.03 percent increase (South submarket).

6.2 Marginal value added

Table 5, column (1) reports summary statistics of the estimated marginal responses obtained using Table 3 and Equation (9). On average, a one unit increase in diversity corresponds to approximately 6.51 additional responses to a home advert. For the majority of adverts the marginal responses are lower, however; the estimated marginal responses follow a right-skewed distribution.

Table 6, column (1) reports summary statistics of the estimated marginal registration length due to increased diversity. On average, a one unit increase in diversity corresponds to approximately seven additional days of registration length. For the majority of homes the marginal registration length is lower, however; the estimated marginal registration lengths follow a right-skewed distribution.

[6] $[\exp(0.0183) - 1] \cdot 100\% = 1.85\%, \ [\exp(0.0287) - 1] \cdot 100\% = 2.91\%$

Table 3: Hedonic Regression Responses

	(1) Centre	(2) North	(3) South	(4) West
Diversity (inv)	0.0240***	0.0183***	0.0299***	0.0287***
	(0.0067)	(0.0047)	(0.0036)	(0.0031)
Log surface area	0.2622	0.8017***	0.7297***	0.8705***
	(0.1701)	(0.1885)	(0.1544)	(0.1712)
Home age	0.0083***	0.0463***	0.0146***	0.0117***
	(0.0012)	(0.0043)	(0.0013)	(0.0034)
WOZ value (inv)	0.0013	0.0077***	−0.0060***	0.0098***
	(0.0011)	(0.0010)	(0.0010)	(0.0021)
Amenities				
Stores and shops	0.1488**	0.2130***	−0.0092	0.0715
	(0.0476)	(0.0609)	(0.0237)	(0.1325)
Health services	−0.1725*	−0.3876***	−0.1433***	−0.0741
	(0.0752)	(0.0646)	(0.0298)	(0.0767)
Hospitality	−0.0666	−0.0556	−0.1923***	0.0634
	(0.1241)	(0.0356)	(0.0417)	(0.0518)
Leasure activities	−0.1573	0.0491	0.2355**	−0.2730
	(0.1042)	(0.0633)	(0.0885)	(0.2080)
Schools	0.1532*	−0.1121***	−0.0776	−0.0748*
	(0.0614)	(0.0299)	(0.0401)	(0.0334)
Public transportation	−0.4947**	−0.3932**	0.5113***	0.1675
	(0.1786)	(0.1439)	(0.1460)	(0.1594)
Neighbourhood safety	−0.0321***	−0.0273***	0.0288***	−0.0074
	(0.0051)	(0.0077)	(0.0035)	(0.0196)
Constant	5.8843***	−0.5900	−2.9760***	−2.7752
	(1.1184)	(1.2295)	(0.8240)	(2.7920)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.175	0.304	0.277	0.341
Observations	922	1050	1560	495

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Hedonic Regression Registration Length

	(1) Centre	(2) North	(3) South	(4) West
Diversity (inv)	0.0101	−0.0046	0.0102**	0.0087**
	(0.0071)	(0.0054)	(0.0034)	(0.0031)
Log surface area	0.4043*	1.6428***	1.3721***	1.2004***
	(0.1748)	(0.2243)	(0.1448)	(0.1826)
Home age	0.0050***	0.0170***	0.0057***	0.0017
	(0.0012)	(0.0050)	(0.0012)	(0.0035)
WOZ value (inv)	0.0031**	0.0052***	−0.0034**	0.0061**
	(0.0011)	(0.0012)	(0.0010)	(0.0019)
Amenities				
Stores and shops	0.0586	−0.0027	−0.0197	0.0170
	(0.0485)	(0.0731)	(0.0252)	(0.1362)
Health services	0.1366*	−0.0667	−0.0541	0.1174
	(0.0646)	(0.0664)	(0.0283)	(0.0735)
Hospitality	0.1163	0.0338	−0.0291	0.1357*
	(0.1204)	(0.0344)	(0.0435)	(0.0554)
Leasure activities	−0.2747**	0.0447	0.2144*	−0.1762
	(0.0991)	(0.0661)	(0.0889)	(0.2296)
Schools	0.0989	−0.0153	0.0453	−0.0074
	(0.0646)	(0.0293)	(0.0396)	(0.0328)
Public transportation	−0.6052***	−0.1640	0.0568	0.0091
	(0.1622)	(0.1597)	(0.1388)	(0.1726)
Neighbourhood safety	−0.0178***	−0.0151*	0.0066	0.0166
	(0.0047)	(0.0073)	(0.0038)	(0.0204)
Constant	6.0304***	−0.2300	0.0946	−2.2396
	(1.1403)	(1.3092)	(0.8541)	(2.8952)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.264	0.112	0.293	0.300
Observations	605	791	1115	449

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Estimated Marginal Responses due to Less Segregation

	(1) Diversity (inv)	(2) Representation (inv)	(3) Diversity (avg)
Mean	6.51	4.63	11.31
SD	5.52	6.08	10.95
Min	0.11	−7.16	0.03
Max	31.63	29.98	53.51

Notes: All statistics reported as amount of responses to an advert.

Table 6: Estimated Marginal Registration Length due to Less Segregation

	(1) Diversity (inv)	(2) Representation (inv)	(3) Diversity (avg)
Mean	7.13	7.12	9.34
SD	8.22	9.64	14.96
Min	−9.62	−9.83	−36.75
Max	37.29	44.44	59.98

Notes: All statistics reported in number of days.

Although the marginal responses are positive for all observations, some observations show negative marginal registration length. This suggests that although all applicants appear to prefer less segregation, not all applicants are willing to "spend" their registration length for this. It could very well be that if applicants obtain additional registration length, they rather make use of this to obtain a home with, for example, a larger surface area.

Similarly, marginal registration length relative to total registration length ($7.13/1187.40 = 0.006$ on average) is much smaller than marginal responses relative to total responses ($6.51/440.18 = 0.015$). These results further suggest that although applicants appear to prefer less segregation, willingness to pay seems small to nil.

6.3 Preference model

Table 7, column (1) reports the estimated coefficients of the preference model (Equation (10)) using the IHS of estimated marginal responses as the dependent variable. All considered socio-economic indicators, except welfare usage, appear to have a significant association with the marginal value of diversity in terms of responses. In terms of age, diversity appears to be valued less for populations with a larger share of elderly residents (-1.66% per percentage point), while it is valued more when children make up a larger share of the population (1.98% per percentage point). At the same time, average household size has a negative association with the marginal value of diversity (-18.19% per household member), counteracting the effect of a younger population share. Both household income and the index for welfare usage have a negative coefficient, suggesting that diversity is valued less both as income increases (-0.47% per 1% increase) as well as when more residents receive welfare (-1.03% per unit increase). Diversity is also valued less as a higher share of the population reports to have difficulties with the Dutch language (-2.78% per percentage point). Finally, diversity is valued more when a larger share of the working age population has a lower education level (1.35% per percentage point).

Table 8, column (1) reports the estimated coefficients of the preference model using the IHS of estimated marginal registration length as the dependent variable. In contrast with

Table 7: Effect of Socio-Economic Status on IHS of Marginal Responses

	(1) Diversity (inv)	(2) Representation (inv)	(3) Diversity (avg)
Share >65 (%)[1]	−0.0167***	−0.0152***	−0.0196***
	(0.0013)	(0.0014)	(0.0017)
Share 0–14 (%)[1]	0.0196***	0.0134**	0.0392***
	(0.0040)	(0.0043)	(0.0059)
Household size[1]	−0.2008***	−0.3457***	−0.4600***
	(0.0575)	(0.0642)	(0.0915)
Log household income[1]	−0.4743***	−0.3480**	−0.1019
	(0.1020)	(0.1117)	(0.1619)
Welfare usage	−0.0104	−0.1228***	0.0249
	(0.0151)	(0.0172)	(0.0150)
Share w/ language barrier (%)	−0.0282***	−0.0519***	−0.0298***
	(0.0025)	(0.0028)	(0.0026)
Share w/o starting qualification (%)	0.0134***	0.0284***	0.0102**
	(0.0035)	(0.0041)	(0.0032)
Constant	7.8413***	2.3738*	5.4635***
	(1.0070)	(1.0951)	(1.6081)
District FE	Yes	Yes	Yes
Allocation dummies	Yes	Yes	Yes
Adj. R^2	0.672	0.913	0.865
Observations	3858	3848	3847

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[1] weighted average in columns (1) and (2), average in column (3)

Table 8: Effect of Socio-Economic Status on IHS of Marginal Registration Length

	(1) Diversity (inv)	(2) Representation (inv)	(3) Diversity (avg)
Share >65 (%)[1]	−0.0107***	−0.0026	−0.0146***
	(0.0014)	(0.0017)	(0.0023)
Share 0–14 (%)[1]	−0.0048	0.0015	−0.0095
	(0.0041)	(0.0048)	(0.0070)
Household size[1]	−0.1501*	0.0884	−0.3056**
	(0.0590)	(0.0679)	(0.1068)
Log household income[1]	−0.4762***	−0.1163	−0.4801*
	(0.1165)	(0.1401)	(0.2050)
Welfare usage	−0.0223	−0.1161***	0.0036
	(0.0155)	(0.0194)	(0.0161)
Share w/ language barrier (%)	−0.0049	−0.0260***	−0.0066*
	(0.0026)	(0.0035)	(0.0029)
Share w/o starting qualification (%)	0.0154***	0.0236***	0.0112**
	(0.0037)	(0.0049)	(0.0036)
Constant	8.1262***	−0.7421	9.3288***
	(1.1456)	(1.3618)	(2.0396)
District FE	Yes	Yes	Yes
Allocation dummies	Yes	Yes	Yes
Adj. R^2	0.957	0.940	0.975
Observations	2853	2850	2850

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[1] weighted average in columns (1) and (2), average in column (3)

marginal responses, less of the considered socio-economic indicators have a significant association with estimated marginal registration length. A larger share of elderly residents again shows a negative relationship, but the share of children does not appear to have a significant effect. In terms of age, willingness to pay for diversity appears lower as the elderly population increases (-1.06% per percentage point), while unaffected by the share of children in the population. As with marginal responses, household size and income show a significant and negative relationship, indicating diversity is valued less as household size and income increase (-13.94% per household member and -0.48% per 1% increase). Compared to the responses model, welfare usage has a larger negative effect (-2.21% per unit increase), while the share of population with a language barrier had a smaller, now insignificant effect (-0.49% per percentage point) on the appreciation of diversity. Finally, the share of the working age population with a lower education level has a similarly sized positive effect (1.55% per percentage point).

Comparing the two models, the most notable differences appear to be in the effect of the share of the population younger than fourteen and with a language barrier. This suggests that although having children or experiencing difficulties with the Dutch language appears affect tenants' preferences for diversity, it nevertheless does not affect their willingness to actually pay for more diversity in their neighbourhood. Furthermore, although household size has a significant and positive effect in both models, it appears to have a larger – and more significant – effect on preferences than willingness to pay. In all, the results of the main models suggest that the key socio-economic factors that influence segregation preferences are old age, household income, education level, and – to a lesser extent – household size.

7 Robustness Checks

This section reports the results of two robustness checks. The first robustness check uses a different definition of segregation to see whether the specific definition used in the main model drives the found results. The second robustness check uses different weights for disaggregation to whether the disaggregation weights drive the found results.

7.1 Representation

The models in this subsection measure segregation as deviation from (citywide) population proportions, as per the second paragraph of Section 4.2, rather than segregation as the opposite of diversity, as in the main model. The key difference in the two definitions is the point of reference used for determining segregation. Using the diversity definition, segregation is lowest when all groups make up an equal share of the population in the area considered. On the other hand, following the disproportionality definition, segregation is lowest when each group makes up a share of the area's population equal to the share that group makes up of the reference (total) population. The results of this robustness check should thus be interpreted slightly differently. In contrast with the main model, a preference for less segregation here means a preference for a group that perfectly represents the reference population, rather than a group that is as (ethnically) diverse as possible.

Hedonic model Table 9 reports the estimated coefficients of the adjusted hedonic regression model, using the amount of responses for an advert as the dependent variable. Similar to the main model, representation has a significant relationship with the amount of responses in each submarket. For the Centre submarket, however, the sign of the coefficient has flipped and now shows a negative relationship. Furthermore, the size of this coefficient has also decreased, suggesting a relatively small preference for more segregation in this submarket of 0.83 percent decrease in responses per unit increase in representation. The other submarkets nevertheless show a positive relationship ranging from 2.70 to 3.59 percent increase in responses per unit increase in representation. All other coefficients are mostly unchanged with respect to the main model, suggesting the control variables in the first stage are mostly unaffected by the segregation definition used in the first stage responses model.

Table 10 reports the estimated coefficients of the same adjusted model using registration length of the succesful applicant as the dependent variable. Again, results are similar to the main model, where the Centre and North submarkets show a statistically insignificant relationship between registration length and segregation, while the South and Western submarket show a positive and significant relationship between registration length and less segregation. The main difference here is that the coefficient sign for representation has flipped for both the Centre and North submarkets, although both remain insignificant.

Marginal value added Table 5, column (2) reports summary statistics of the estimated marginal responses obtained using the adjusted model. On average, a one unit increase in representation corresponds to approximately 4.63 addition responses to an advert. Although the mean is lower than in the main model, the values again follow a right-skewed distribution. Furthermore, the range of the values is larger due to the negative lower bound.

Table 6, column (2) reports summary statistics of the estimated marginal registration length obtained using the adjusted model. As with the responses model, the standard deviation is slightly larger than the main model. The range of values is also larger than in the main model. In contrast with the responses model, here this si the case due to a shift in the upper bound, which is further away from zero.

Table 9: Hedonic Regression Responses

	(1) Centre	(2) North	(3) South	(4) West
Representation (inv)	−0.0083*	0.0311***	0.0266***	0.0353***
	(0.0042)	(0.0046)	(0.0039)	(0.0037)
Log surface area	0.2754	0.7807***	0.7056***	0.8880***
	(0.1708)	(0.1869)	(0.1575)	(0.1720)
Home age	0.0088***	0.0474***	0.0143***	0.0116***
	(0.0012)	(0.0045)	(0.0013)	(0.0034)
WOZ value (inv)	0.0019	0.0071***	−0.0075***	0.0060**
	(0.0011)	(0.0010)	(0.0010)	(0.0020)
Amenities				
Stores and shops	0.1365**	0.1872**	0.0431	0.0280
	(0.0483)	(0.0581)	(0.0264)	(0.1323)
Health services	−0.1810*	−0.3506***	−0.0910**	−0.0683
	(0.0760)	(0.0641)	(0.0282)	(0.0761)
Hospitality	0.0082	−0.0568	−0.1706***	0.1126*
	(0.1305)	(0.0351)	(0.0417)	(0.0519)
Leasure activities	−0.1772	0.0897	0.0926	−0.3223
	(0.1057)	(0.0639)	(0.0890)	(0.2087)
Schools	0.1708**	−0.1294***	−0.1166**	−0.1189***
	(0.0610)	(0.0307)	(0.0405)	(0.0333)
Public transportation	−0.5536**	−0.3400*	0.5319***	0.2592
	(0.1840)	(0.1456)	(0.1557)	(0.1627)
Neighbourhood safety	−0.0347***	−0.0248**	0.0242***	−0.0208
	(0.0052)	(0.0075)	(0.0035)	(0.0194)
Constant	8.5171***	−2.2150	−2.0611**	−1.4442
	(0.9697)	(1.2009)	(0.7978)	(2.7322)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.165	0.314	0.272	0.340
Observations	922	1041	1555	495

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Hedonic Regression Registration Length

	(1) Centre	(2) North	(3) South	(4) West
Representation (inv)	−0.0030	0.0033	0.0121***	0.0126**
	(0.0047)	(0.0060)	(0.0033)	(0.0042)
Log surface area	0.3943*	1.5915***	1.3540***	1.1805***
	(0.1758)	(0.2297)	(0.1465)	(0.1819)
Home age	0.0051***	0.0161**	0.0053***	0.0016
	(0.0012)	(0.0050)	(0.0012)	(0.0035)
WOZ value (inv)	0.0032**	0.0052***	−0.0039***	0.0052**
	(0.0011)	(0.0012)	(0.0010)	(0.0019)
Amenities				
Stores and shops	0.0537	0.0179	−0.0003	0.0101
	(0.0490)	(0.0716)	(0.0266)	(0.1341)
Health services	0.1345*	−0.0669	−0.0390	0.1086
	(0.0655)	(0.0687)	(0.0274)	(0.0730)
Hospitality	0.1495	0.0216	−0.0122	0.1490**
	(0.1226)	(0.0345)	(0.0433)	(0.0551)
Leasure activities	−0.2930**	0.0409	0.1764*	−0.2134
	(0.0986)	(0.0660)	(0.0875)	(0.2281)
Schools	0.1089	−0.0307	0.0294	−0.0256
	(0.0629)	(0.0307)	(0.0405)	(0.0330)
Public transportation	−0.6256***	−0.1382	0.0900	0.0586
	(0.1632)	(0.1620)	(0.1413)	(0.1726)
Neighbourhood safety	−0.0187***	−0.0120	0.0056	0.0094
	(0.0047)	(0.0071)	(0.0038)	(0.0202)
Constant	7.1823***	−0.9136	0.1172	−1.5133
	(0.9785)	(1.2848)	(0.8262)	(2.8360)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.261	0.111	0.294	0.305
Observations	605	791	1110	449

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In both cases, the values appear more spread out than in main model, but remain mostly similar. Of particular note is the change in the lower bound for the responses model, which no longer shows only positive values as was the case in the main model. This suggests that, in contrast with the main model, there might not be a strict preference for less segregation in certain areas.

Preference model Table 7, column (2) reports the estimated coefficients of the adjusted preference model using the IHS of marginal responses as the dependent variable. All coefficients show the same sign as in the main model, yet most differ in size. Of particular note is welfare usage, which is now significant and has a ten times larger effect size. As most coefficients show are larger than the main model, it appears that preferences for better representation depend more on the socio-economic status of a neighbourhood's current residents compared to preferences for more diversity.

Table 8, column (2) reports the estimated coefficients of the adjusted preference model using the IHS of marginal registration length as the dependent variable. In contrast with the responses models, most of the coefficients of the registration length model have changed with respect to the main model. The coefficients for the share of elderly population, household size, and household income are no longer significant. Instead, welfare usage and the share of current residents experiencing a language barrier now have a significant and negative effect on willingness to pay for better representation. As in the main model, the share of current residents with a lower education level has a significant and positive effect on estimated marginal registration length. It appears willingness to pay for more representation is affected by different socio-economic indicators than willingness to pay for more diversity. Although estimated marginal registration length appears similar between the two definitions of segregation, the preferences model shows that the underlying mechanism is not robust to the used definition.

7.2 Variable disaggregation

The models in this subsection make use of variables disaggregated with a non-weighted average, rather than a weighted average as in the main model. As with the main model, segregation is defined as the inverse of diversity. The results of this robustness check should show whether the results obtained in the main model are sensitive to the weights used. Appendix C reports an additional robustness check using non-disaggregated variables.

Hedonic model Table 11 reports the estimated coefficients of the adjusted hedonic regression model, using the amount of responses for an advert as the dependent variable. As in the main model, the coefficient for diversity is positive in all submarkets, although in contrast to the main model it is not significant in the North submarket. In the other three submarkets, the diversity coefficient is much larger compared to the main model,

Table 11: Hedonic Regression Responses

	(1) Centre	(2) North	(3) South	(4) West
Diversity (avg)	0.0647***	0.0072	0.0481***	0.0411***
	(0.0120)	(0.0061)	(0.0055)	(0.0047)
Log surface area	0.1435	1.0371***	0.7372***	0.8702***
	(0.1653)	(0.1920)	(0.1535)	(0.1762)
Home age	0.0082***	0.0471***	0.0157***	0.0132***
	(0.0012)	(0.0045)	(0.0013)	(0.0034)
WOZ value (avg)	0.0054***	0.0091***	−0.0072***	0.0177***
	(0.0015)	(0.0013)	(0.0013)	(0.0038)
Amenities				
Stores and shops	0.1528**	0.2620***	−0.0243	−0.0159
	(0.0486)	(0.0629)	(0.0237)	(0.1352)
Health services	−0.1648*	−0.4322***	−0.1774***	−0.0434
	(0.0752)	(0.0675)	(0.0316)	(0.0795)
Hospitality	−0.1388	−0.0506	−0.1902***	0.0735
	(0.1241)	(0.0356)	(0.0421)	(0.0549)
Leasure activities	−0.1516	0.0968	0.2887**	−0.1792
	(0.1027)	(0.0644)	(0.0893)	(0.2066)
Schools	0.1373*	−0.0857**	−0.0916*	−0.0368
	(0.0609)	(0.0305)	(0.0403)	(0.0384)
Public transportation	−0.5420**	−0.6041***	0.5713***	−0.0839
	(0.1823)	(0.1575)	(0.1482)	(0.1539)
Neighbourhood safety	−0.0263***	−0.0323***	0.0357***	0.0252
	(0.0054)	(0.0086)	(0.0038)	(0.0199)
Constant	1.4543	−0.6272	−5.1338***	−10.3204***
	(1.5261)	(1.3822)	(0.9358)	(2.9637)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.199	0.288	0.279	0.329
Observations	922	1041	1554	495

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Hedonic Regression Registration Length

	(1) Centre	(2) North	(3) South	(4) West
Diversity (avg)	0.0179	−0.0165*	0.0167**	0.0091
	(0.0129)	(0.0075)	(0.0052)	(0.0048)
Log surface area	0.3980*	1.7253***	1.3828***	1.2416***
	(0.1791)	(0.2217)	(0.1460)	(0.1851)
Home age	0.0048***	0.0171***	0.0059***	0.0020
	(0.0012)	(0.0048)	(0.0012)	(0.0035)
WOZ value (avg)	0.0040**	0.0063***	−0.0044**	0.0101**
	(0.0014)	(0.0015)	(0.0014)	(0.0032)
Amenities				
Stores and shops	0.0646	0.0453	−0.0312	−0.0413
	(0.0497)	(0.0764)	(0.0256)	(0.1413)
Health services	0.1375*	−0.0945	−0.0629*	0.1450
	(0.0647)	(0.0676)	(0.0296)	(0.0747)
Hospitality	0.1000	0.0280	−0.0251	0.1557**
	(0.1200)	(0.0334)	(0.0434)	(0.0577)
Leasure activities	−0.2558**	0.0747	0.2308*	−0.1267
	(0.0974)	(0.0677)	(0.0904)	(0.2318)
Schools	0.0939	0.0082	0.0446	0.0193
	(0.0644)	(0.0295)	(0.0399)	(0.0344)
Public transportation	−0.6277***	−0.3443	0.1176	−0.0870
	(0.1678)	(0.1760)	(0.1433)	(0.1693)
Neighbourhood safety	−0.0157**	−0.0249**	0.0100*	0.0325
	(0.0051)	(0.0082)	(0.0040)	(0.0205)
Constant	4.9270**	1.1833	−0.6690	−5.6319
	(1.5276)	(1.5714)	(0.9883)	(2.9881)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.262	0.118	0.293	0.295
Observations	605	791	1110	449

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

suggesting a positive relationship ranging from a 4.20 to 6.68 percent increase in responses per unit increase in diversity. All other coefficients remain largely unchanged, with the exception of home value, which is now significant in all four submarkets and has a larger effect in the Centre and West submarkets.

Table 12 reports the estimated coefficients of the same adjusted model using registration length of the succesful applicant as the dependent variable. In the North submarket, the negative relationship between diversity and registration length has increased and is now statistically significant. The positive relationship in de West submarket, on the other hand, is no longer significant. The relationship remains insignificant in the Centre submarket and significant in the South submarket, with the coefficient being larger than the main model in both submarkets. All other coefficients appear mostly unaffected by the different disaggregation method.

For both adjusted hedonic models it appears that using a non-weighted average increases the size of the statistically significant diversity coefficients, suggesting the diversity of the entire neighbourhood has a stronger effect on the home value than diversity of the area directly surrounding a home.[7]

Marginal value added Table 5, column (3) reports summary statistics of the estimated marginal responses obtained using the adjusted model. On average, a one unit increase in diversity corresponds to approximately 11.31 additional responses to an advert. Again, the values follow a right-skewed distribution. Compared to the main model, the mean, standard deviation, and upper bound are all almost twice as large. This increase matches the increase in the coefficients of the hedonic model, suggesting a stronger preference for diversity than found in the main model.

Table 6, column (3) reports summary statistics of the estimated marginal registration length obtained using the adjusted model. On average, a one unit increase in diversity corresponds to approximately nine additional days of registration length. In contrast with the main model and the estimated marginal responses, the values do not appear to follow a right-skewed distribution. Furthermore, compared to the main model both upper and lower bound are much further away from zero, and the standard deviation is almost twice as large. Compared to the main model, willingness to pay for less segregation varies more and values are more likely to be negative.

Preference model Table 7, column (3) reports the estimated coefficients of the adjusted preference model using the IHS of marginal responses as the dependent variable. Most coefficients show the same sign as in the main model, except for welfare usage, which –

[7] Recall that the main model puts more weight on smaller areas, so that the diversity of the area directly surrounding a home gets a higher weight than the diversity of the entire neighbourhood the home is located in. As this robustness check uses equal weights, the values for larger areas gain a higher weight compared to the main model.

although statistically insignificant – has become positive. The coefficients for the share of elderly population, the share of residents experiencing difficulties with the Dutch language, and the share of residents with a lower education level have remained statistically significant and approximately similar in size. The coefficients for the share of children and for household size, on the other hand, have at least doubled in size compared to the main model. Finally, the coefficient for household income has become smaller and is now statistically insignificant, suggesting household income in an area might not have an effect on preferences for less segregation of (potential) new residents.

Table 8, column (3) reports the estimated coefficients of the adjusted preference model using the IHS of marginal registration length as the dependent variable. Compared to the adjusted responses model, the coefficients show less differences with respect to the main model. The coefficients for household size and the share of residents experiencing a language barrier show the most notable changes. Compared to the main model, the effect of household size has approximately doubled. On the other hand, the effect of the share of residents experiencing a language barrier has not increased as much, but is now statistically significant. It appears that the different set of weights used for disaggregation does not have a strong effect on which socio-economic indicators influence preferences and willingness to pay for more diversity, suggesting the results of the preferences model are at least somewhat robust to the choice of weights for disaggregation.

8 Conclusion

This thesis examines how the socio-economic status of an area's residents influences a (potential) new resident's preferences for segregation in Rotterdam social housing. Using a modified hedonic pricing model with novel data on social housing in Rotterdam, I found that homes in less segregated areas are, in general, preferred over homes in more segregated areas. In spite of this, willingness to pay for less segregation appears to be low to nonexistent. Furthermore, I find that there are several socio-economic aspects of the neighbourhood's current residence that affect the desirability of less segregation for that neighbourhood, namely the share of elderly residents, the average household size, household income, and the share of residents with a lower education level. The share of children in a neighbourhood and the share of residents that experience difficulties with the Dutch language also appear to affect preferences for less segregation, but do not show a significant effect on willingness to pay.

These results add to the existing literature in two distinct ways. First, current literature examining the relationship between socio-economic status and segregation preferences focuses on the socio-economic status of a prospective resident. To the best of my knowledge the effect of socio-economic status of *current* residents on the desirability of segregation has thus far not been examined in the literature. The results in this thesis suggest, however, that socio-economic status of current residents also determines the extent to which less segregation is perceived as desirable. Second, preceding literature mostly investigates the home-owners' market, and studies that investigate the rental market appear to examine only the "unregulated" rental market. This thesis, on the other hand, solely examines social (rental) housing, thereby targeting a sub-population generally perceived to have a very limited choice set with respect to housing. Although this sub-population has limited means to pursue personal preferences, I find evidence that (a lack of) segregation affects the desirability of a home, even in this restricted market.

Based on my results, policymakers might consider an indirect approach when aiming to reduce segregation. Rather than directly reducing ethnic segregation by, for example, giving certain prospective residents preferential treatment, they might instead target policy at other socio-economic indicators of that neighbourhood, so that diversity ends up in a more positive daylight, reducing segregation through a second order effect as potential new residents themselves look for less segregated living areas.

Finally, some caveats of this thesis should be addressed. Because the data on Rotterdam social housing was collected by the author, only a single year of observations was available, restricting this thesis to a cross-sectional analysis. Furthermore, no socio-economic data of new residents and only aggregated socio-economic geographic data was available, precluding analysis based on characteristics of prospective residents and limiting statistical power. Further research might thus extend this thesis using more granular data to examine the interplay between neighbourhood socio-economic status, prospective resident socio-economic status, and segregation preferences, or use panel- or repeated cross-sectional data to (attempt to) estimate a causal relationship.

References

Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models. *Journal of Political Economy*, *95*(1), 81–88. https://doi.org/10.1086/261442

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, *82*(1), 50–61. https://doi.org/10.1111/obes.12325

Brown, J. N., & Rosen, H. S. (1982). On the estimation of structural hedonic price models. *Econometrica*, *50*(3), 765–768. https://doi.org/10.2307/1912614

Clark, W. A. (2009). Changing residential preferences across income, education, and age: Findings from the multi-city study of urban inequality. *Urban Affairs Review*, *44*(3), 334–355. https://doi.org/10.1177/1078087408321497

Daniels, C. B. (1975). The influence of racial segregation on housing prices. *Journal of Urban Economics*, *2*(2), 105–122. https://doi.org/10.1016/0094-1190(75)90042-X

Doucet, B., & Koenders, D. (2018). 'At least it's not a ghetto anymore': Experiencing gentrification and 'false choice urbanism' in Rotterdam's Afrikaanderwijk. *Urban Studies*, *55*(16), 3631–3649. https://doi.org/10.1177/0042098018761853

Gemeente Rotterdam. (2020). *Rotterdam tegen Racisme*. https://rotterdam.notubiz.nl/document/9331139/1

Hwang, S.-S., Murdock, S. H., Parpia, B., & Hamm, R. R. (1985). The effects of race and socioeconomic status on residential segregation in texas, 1970-80. *Social Forces*, *63*(3), 732–747. https://doi.org/10.2307/2578489

Ibraimovic, T., & Hess, S. (2018). A latent class model of residential choice behaviour and ethnic segregation preferences. *Housing Studies*, *33*(4), 544–564. https://doi.org/10.1080/02673037.2017.1373749

Ibraimovic, T., & Masiero, L. (2014). Do birds of a feather flock together? The impact of ethnic segregation preferences on neighbourhood choice. *Urban Studies*, *51*(4), 693–711. https://doi.org/10.1177/0042098013493026

Iceland, J., & Wilkes, R. (2006). Does socioeconomic status matter? race, class, and residential segregation. *Social problems*, *53*(2), 248–273. https://doi.org/10.1525/sp.2006.53.2.248

Järv, O., Masso, A., Silm, S., & Ahas, R. (2021). The link between ethnic segregation and socio-economic status: An activity space approach. *Journal of Economic and Human Geography*, *112*(3), 319–335. https://doi.org/https://doi.org/10.1111/tesg.12465

Mei, Y., Hite, D., & Sohngen, B. (2017). Demand for urban tree cover: A two-stage hedonic price analysis in California. *Forest Policy and Economics*, *83*, 29–35. https://doi.org/10.1016/j.forpol.2017.05.009

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing. *The Review of Economics and Statistics*, *66*(3), 394. https://doi.org/10.2307/1924995

Poudyal, N. C., Hodges, D. G., & Merrett, C. D. (2009). A hedonic analysis of the demand for and benefits of urban recreation parks. *Land Use Policy*, *26*(4), 975–983. https://doi.org/10.1016/j.landusepol.2008.11.008

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological methodology*, *32*(1), 33–67. https://doi.org/10.1111/1467-9531.00110

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55. https://doi.org/10.1086/260169

Simons, R., Karremans, V., Zeegers, C., & Achbar, F. (2022). *Eén Stad: Coalitieakkoord 2022-2026 Rotterdam*. https://openrotterdam.nl/wp-content/uploads/2022/06/Coalitieakkoord-2022-2026_Een-Stad_Rotterdam.pdf

Spivak, A. L., & Monnat, S. M. (2013). The influence of race, class, and metropolitan area characteristics on african-american residential segregation. *Social Science Quarterly*, *94*(5), 1414–1437. https://doi.org/10.1111/ssqu.12021

Verdugo, G., & Toma, S. (2018). Can public housing decrease segregation? Lessons and challenges from non-European immigration in France: Can public housing decrease segregation? *Demography*, *55*(5), 1803–1828. https://doi.org/10.1007/s13524-018-0705-4

Zhang, J., & Zheng, L. (2015). Are people willing to pay for less segregation? Evidence from US internal migration. *Regional Science and Urban Economics*, *53*, 97–112. https://doi.org/10.1016/j.regsciurbeco.2015.05.002

A Observation Levels

Figure A1: Observation Levels

Notes: Map depicting different levels of observation for illustrative purposes. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

B Indices Used

Table B1: PCA Loadings for Stores Index

	Weight
Closest ... in km	
Supermarket	−0.2030
Store	−0.3208
Warehouse	−0.1371
Share of homes with ... within norm distance	
Baker	0.4570
Greengrocer	0.4583
Butcher	0.4570
Drugstore	0.4576

Notes: weights are set equal to the loadings of the first principal component of the variables.

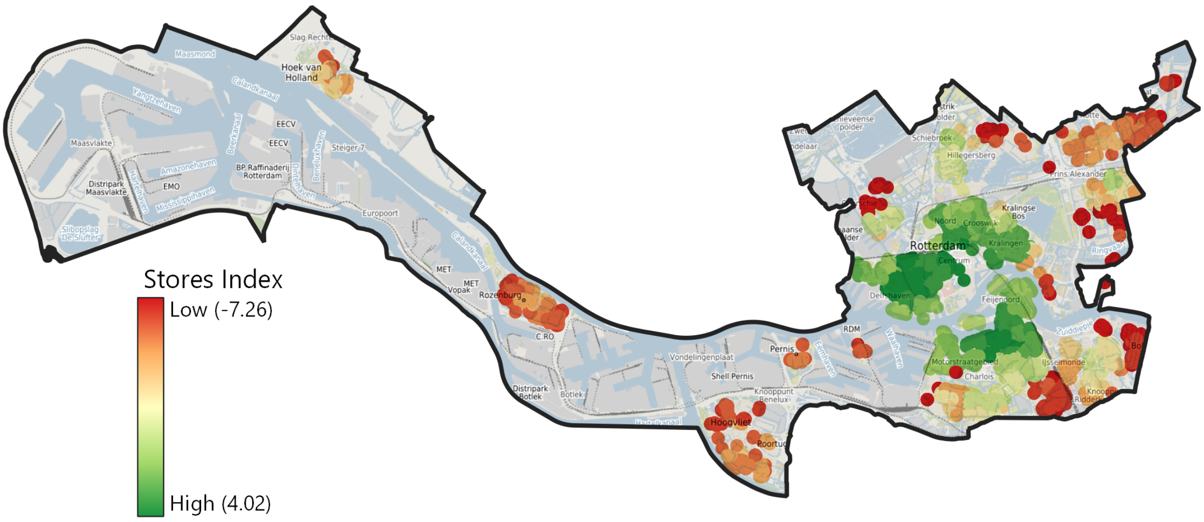

Figure B1: Stores Index Distribution

Notes: Map depicting distribution of the index measuring availability of stores. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table B2: PCA Loadings for Health Services Index

	Weight
Closest ... in km	
General practicioner	−0.3226
Hospital	−0.3087
Outdoor clinic	−0.2445
Pharmacy	−0.2463
General practice centre	−0.2920
Share of homes with ... within norm distance	
Physiotherapist	0.3384
Dentist	0.3944
General practictioner	0.4240
Pharmacy	0.3809

Notes: weights are set equal to the loadings of the first principal component of the variables.

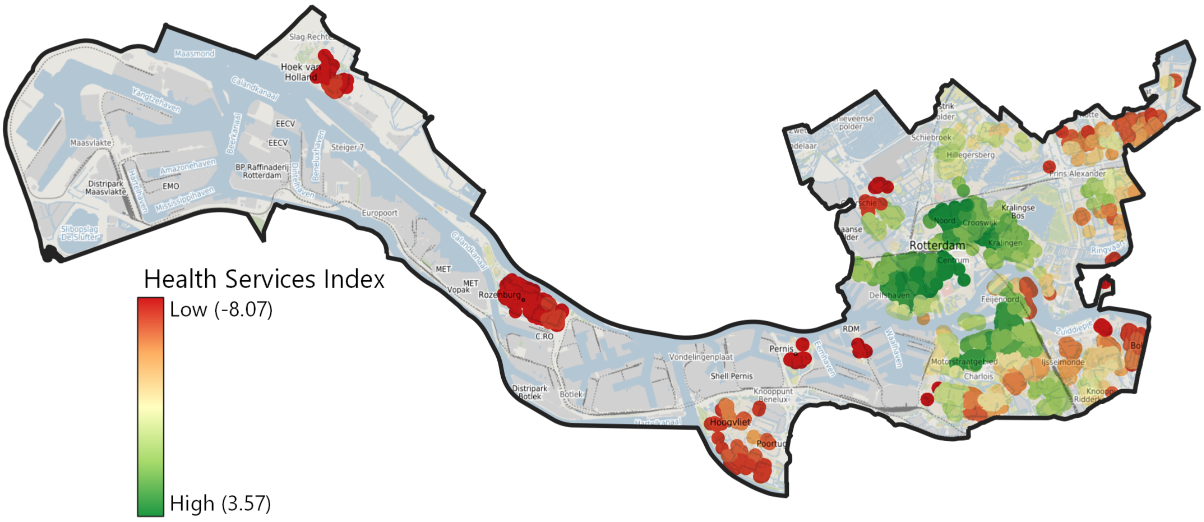

Figure B2: Health Services Index Distribution

Notes: Map depicting distribution of the index measuring availability of health services. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table B3: PCA Loadings for Hospitality Index

	Weight
Closest ... in km	
Cafe	−0.4900
Cafetaria	−0.5376
Restaurant	−0.5256
Hotel	−0.4411

Notes: weights are set equal to the loadings of the first principal component of the variables.

Figure B3: Hospitality Index Distribution

Notes: Map depicting distribution of the index measuring availability of hospitality services. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table B4: PCA Loadings for Leasure Activities Index

	Weight
Closest ... in km	
Cinema	−0.4554
Museum	−0.3725
Theater	−0.4663
Music venue	−0.4179
Share of homes with ... within norm distance	
Sports hall	0.0335
Swimming pool	0.0233
Playground	0.3740
Gym	0.3466

Notes: weights are set equal to the loadings of the first principal component of the variables.

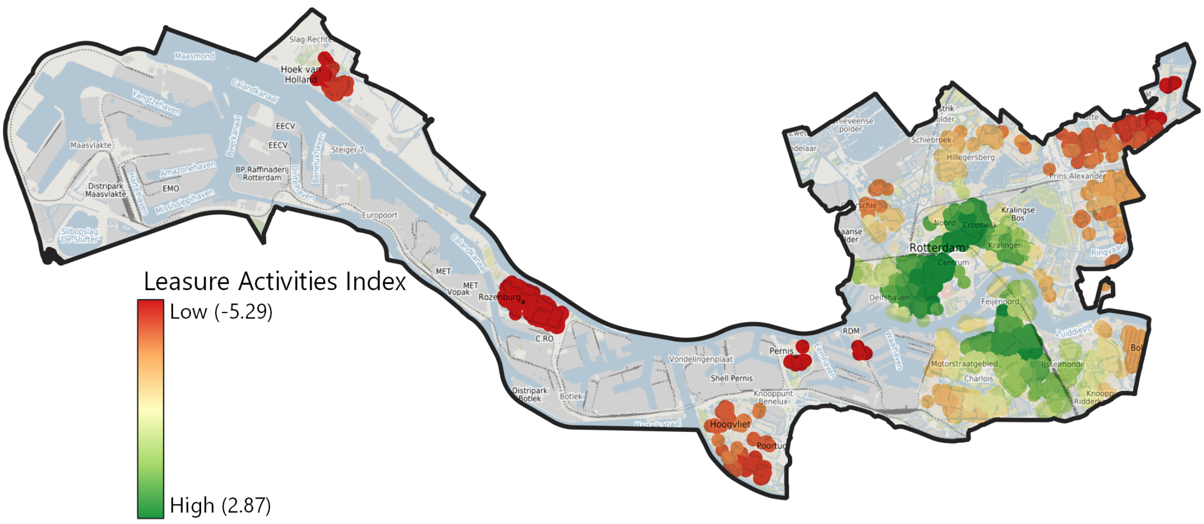

Figure B4: Leasure Activities Index Distribution
Notes: Map depicting distribution of the index measuring availability of leasure activities. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table B5: PCA Loadings for School Index

	Weight
Closest … in km	
School care	−0.0930
Daycare	−0.1653
Primary school	−0.1990
HAVO/VWO school	−0.4107
VMBO school	−0.4384
Secondary school	−0.4390
Share of homes with … within norm distance	
Primary school	0.2372
VMBO school	0.4277
HAVO/VWO school	0.3628

Notes: weights are set equal to the loadings of the first principal component of the variables.

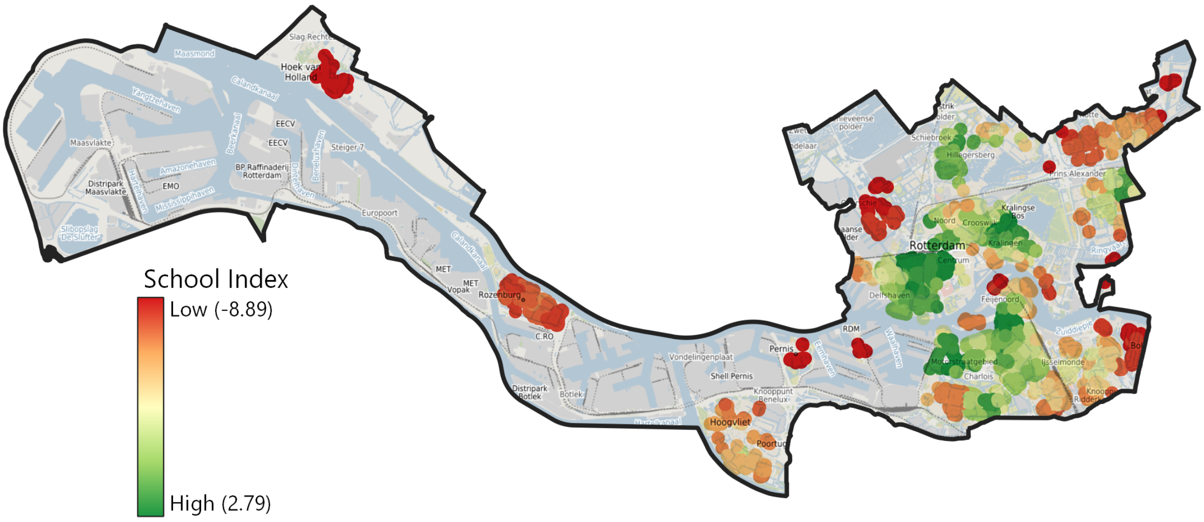

Figure B5: School Index Distribution

Notes: Map depicting distribution of the index measuring availability of schools. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table B6: PCA Loadings for Public Transport Index

	Weight
Closest … in km	
Transfer station	−0.5422
Train station	−0.5491
Share of homes with … within norm distance	
Bus stop	0.0967
Metro station	0.0203
Tram stop	0.4299
Share of residents who think there is enough public transport	0.4581

Notes: weights are set equal to the loadings of the first principal component of the variables.

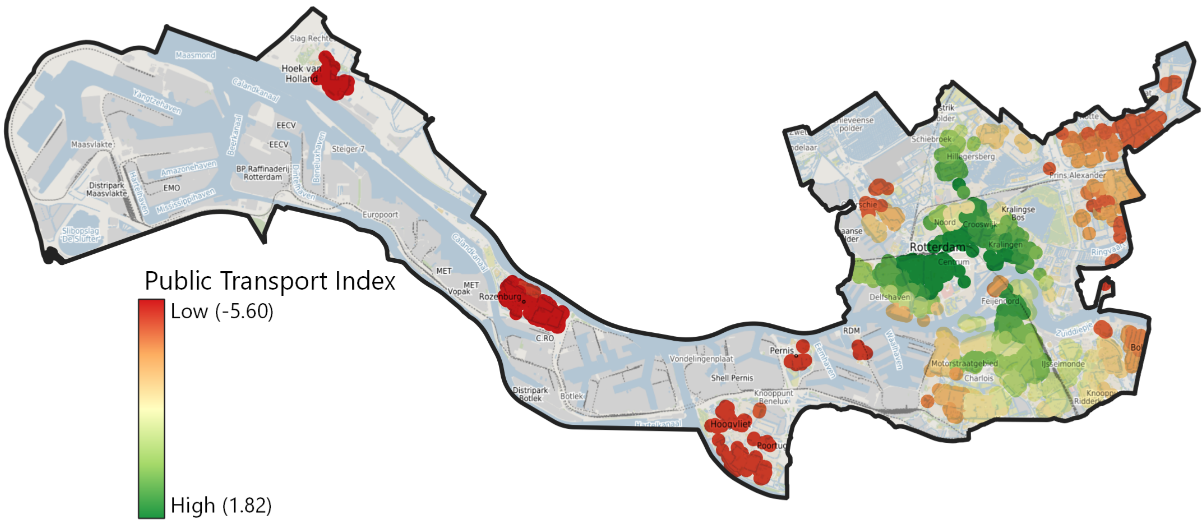

Figure B6: Public Transport Index Distribution

Notes: Map depicting distribution of the index measuring availability of public transportation. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

Table B7: PCA Loadings for Welfare Usage Index

	Weight
Share of neighbourhood with ...	
WIA	0.4799
Bijstand	0.4741
No welfare	−0.4971
Share of PC5 area with welfare	0.3531
Share of 500m^2 area with welfare	0.4161

Notes: weights are set equal to the loadings of the first principal component of the variables.

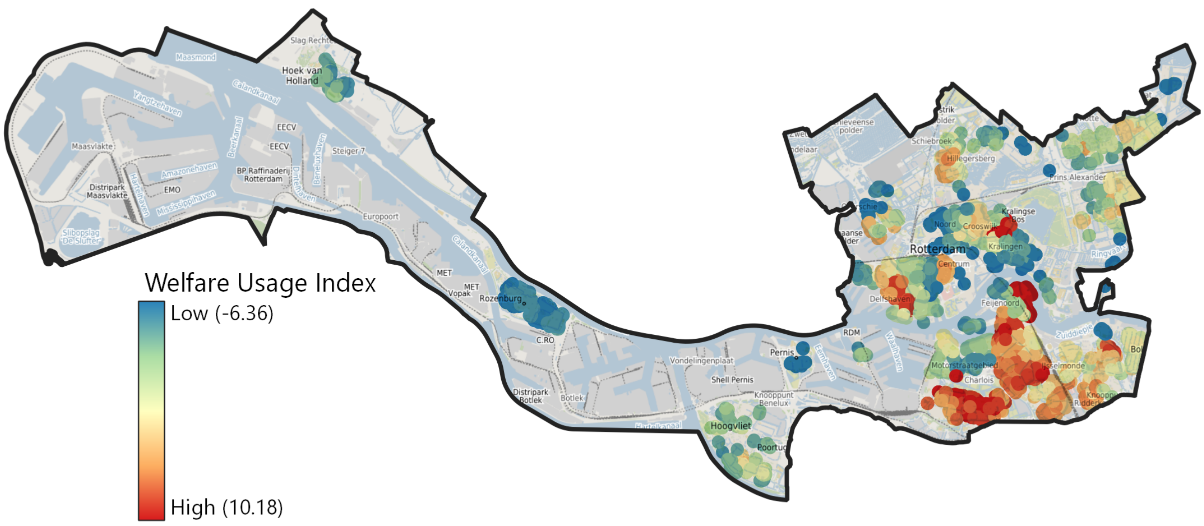

Figure B7: Welfare Usage Index Distribution

Notes: Map depicting distribution of the index measuring welfare usage. Each circle represents one observation. *Source*: Basemap from https://openbasiskaart.nl/, edits by author.

C Robustness Check: No Disaggregation

The tables below show the results of the main model re-estimated without disaggregating variables, using variables on the 5 digit postal code level instead. Tables C3 to C6 include results of the main model in the first column for comparison.

Table C1: Hedonic Regression Responses

	(1) Centre	(2) North	(3) South	(4) West
Diversity	0.0198***	0.0181***	0.0243***	0.0238***
	(0.0056)	(0.0040)	(0.0033)	(0.0024)
Log surface area	0.2766	0.7751***	0.7133***	0.8143***
	(0.1712)	(0.1900)	(0.1558)	(0.1624)
Home age	0.0082***	0.0462***	0.0154***	0.0110**
	(0.0012)	(0.0043)	(0.0013)	(0.0034)
WOZ value	0.0006	0.0073***	−0.0033***	0.0092***
	(0.0009)	(0.0010)	(0.0007)	(0.0016)
Amenities				
Stores and shops	0.1443**	0.1953**	−0.0074	0.2045
	(0.0472)	(0.0596)	(0.0246)	(0.1295)
Health services	−0.1736*	−0.3705***	−0.1274***	−0.0612
	(0.0752)	(0.0635)	(0.0293)	(0.0718)
Hospitality	−0.0566	−0.0538	−0.1926***	0.0617
	(0.1232)	(0.0350)	(0.0431)	(0.0490)
Leasure activities	−0.1526	0.1019	0.1538	−0.4212*
	(0.1043)	(0.0658)	(0.0915)	(0.2083)
Schools	0.1502*	−0.1044***	−0.0382	−0.0796*
	(0.0607)	(0.0298)	(0.0395)	(0.0327)
Public transportation	−0.4763**	−0.4331**	0.3785*	0.2802
	(0.1760)	(0.1434)	(0.1538)	(0.1599)
Neighbourhood safety	−0.0318***	−0.0214**	0.0202***	−0.0183
	(0.0050)	(0.0075)	(0.0034)	(0.0191)
Constant	6.3070***	−0.9622	−2.0689*	−0.4445
	(1.0680)	(1.2224)	(0.8185)	(2.7214)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.174	0.301	0.271	0.365
Observations	922	1041	1543	494

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C2: Hedonic Regression Registration Length

	(1) Centre	(2) North	(3) South	(4) West
Diversity	0.0087	−0.0014	0.0077**	0.0092***
	(0.0058)	(0.0047)	(0.0029)	(0.0025)
Log surface area	0.3939*	1.6013***	1.3406***	1.2056***
	(0.1755)	(0.2289)	(0.1460)	(0.1758)
Home age	0.0050***	0.0149**	0.0058***	0.0010
	(0.0012)	(0.0049)	(0.0012)	(0.0035)
WOZ value	0.0028**	0.0042***	−0.0019*	0.0048**
	(0.0011)	(0.0011)	(0.0008)	(0.0016)
Amenities				
Stores and shops	0.0499	−0.0236	−0.0273	0.0969
	(0.0475)	(0.0726)	(0.0260)	(0.1327)
Health services	0.1368*	−0.0498	−0.0478	0.0802
	(0.0641)	(0.0665)	(0.0280)	(0.0699)
Hospitality	0.1230	0.0266	−0.0275	0.1207*
	(0.1196)	(0.0339)	(0.0442)	(0.0544)
Leasure activities	−0.2891**	0.0538	0.1928*	−0.2276
	(0.1009)	(0.0670)	(0.0888)	(0.2287)
Schools	0.0951	−0.0242	0.0722	−0.0169
	(0.0639)	(0.0290)	(0.0399)	(0.0327)
Public transportation	−0.5863***	−0.0966	−0.0021	0.0606
	(0.1593)	(0.1555)	(0.1424)	(0.1737)
Neighbourhood safety	−0.0180***	−0.0111	0.0033	0.0076
	(0.0047)	(0.0071)	(0.0038)	(0.0202)
Constant	6.3218***	−0.4027	0.4690	−0.7200
	(1.0874)	(1.2867)	(0.8442)	(2.8480)
Allocation dummies	Yes	Yes	Yes	Yes
Adj. R^2	0.263	0.106	0.289	0.310
Observations	605	791	1110	448

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C3: Estimated Marginal Responses due to Less Segregation

	(1) Diversity (inv)	(2) Diversity (pc5)
Mean	6.44	5.43
SD	5.39	4.40
Min	0.11	0.13
Max	30.04	22.72

Notes: All statistics reported as amount of responses to an advert.

Table C4: Estimated Marginal Waiting Time due to Less Segregation

	(1) Diversity (inv)	(2) Diversity (pc5)
Mean	7.12	6.53
SD	8.21	6.82
Min	−9.62	−3.04
Max	37.29	32.32

Notes: All statistics reported in number of days.

Table C5: Effect of Socio-Economic Status on IHS of Marginal Responses

	(1) Diversity (inv)	(2) Diversity (pc5)
Share >65 (%)	−0.0167***	−0.0144***
	(0.0013)	(0.0012)
Share 0–14 (%)	0.0196***	0.0201***
	(0.0040)	(0.0034)
Household size	−0.2008***	−0.2265***
	(0.0575)	(0.0466)
Log household income	−0.4743***	−0.4750***
	(0.1020)	(0.0814)
Welfare usage	−0.0104	−0.0371**
	(0.0151)	(0.0134)
Share w/ language barrier (%)	−0.0282***	−0.0302***
	(0.0025)	(0.0024)
Share w/o starting qualification (%)	0.0134***	0.0195***
	(0.0035)	(0.0031)
Constant	7.8413***	7.4949***
	(1.0070)	(0.8036)
District FE	Yes	Yes
Allocation dummies	Yes	Yes
Adj. R^2	0.672	0.620
Observations	3858	3940

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C6: Effect of Socio-Economic Status on IHS of Marginal Registration Length

	(1) Diversity (inv)	(2) Diversity (pc5)
Share >65 (%)	−0.0107***	−0.0070***
	(0.0014)	(0.0012)
Share 0–14 (%)	−0.0048	0.0033
	(0.0041)	(0.0030)
Household size	−0.1501*	−0.1307***
	(0.0590)	(0.0370)
Log household income	−0.4762***	−0.2557**
	(0.1165)	(0.0821)
Welfare usage	−0.0223	−0.0132
	(0.0155)	(0.0133)
Share w/ language barrier (%)	−0.0049	−0.0031
	(0.0026)	(0.0023)
Share w/o starting qualification (%)	0.0154***	0.0133***
	(0.0037)	(0.0033)
Constant	8.1262***	5.5925***
	(1.1456)	(0.7978)
District FE	Yes	Yes
Allocation dummies	Yes	Yes
Adj. R^2	0.957	0.939
Observations	2853	2922

Notes: Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$