Master's Thesis

Social Housing and Neighbourhood Desirability in Amsterdam: Exploring a Non-Linear Relationship

Author	Ruben Schaap
Student no.	2792369
Supervisor	Dr. Gabriel Loumeau
Date	29 June 2025
Programme	MSc Spatial, Transport and Environmental Economics
Track	Urban and Regional Economics

Abstract

Decisions about the spatial distribution of social housing have far-reaching, unintended external effects. This thesis examines one of these effects by analysing how the local share of social housing relates to neighbouring house prices in Amsterdam, used as a proxy for neighbourhood desirability. It investigates the shape of this relationship; whether it's (non-)linear or even marked by a tipping point. Drawing on over 100,000 observations, the study finds a negative, non-linear association, but no sharp tipping point. This suggests that housing tenure composition matters for neighbourhood outcomes, but in a more nuanced and context-dependent way than some threshold-based theories imply.

Word count: 17,452

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Gabriel Loumeau, for his guidance throughout this project. His critical feedback, clear way of thinking and readiness to question my assumptions helped me improve both the structure and content of the thesis. I especially appreciate his encouragement to explore this topic, which I found genuinely interesting and worthwhile to work on.

I also want to thank the Dutch Association of Real Estate Agents (NVM) for granting me access to the housing transaction data, which formed the empirical backbone of this research. Along those lines, I also acknowledge the Municipality of Amsterdam for making relevant neighbourhood data publicly available.

Thanks go to my fellow students, whose informal feedback, discussions and support played a valuable role during the writing process. Finally, I am grateful to the teaching staff of the Master's programme for providing a strong academic foundation in spatial economics, on which this thesis could build.

Contents

1. Introduction

Across the Netherlands and in many Western countries, housing is at the centre of the political and public debate. This debate is largely dominated by concerns over affordability, shortages and construction-related problems. In response, policies primarily target supply and demand at the national or regional level. For instance, the Dutch government aims to construct 100,000 new homes annually (Dutch National Government, 2024) and Amsterdam applies the 40-40-20 rule to regulate tenure composition in new developments (Municipality of Amsterdam, 2025). These policies focus on aggregate outcomes but pay less attention to the local environments that are also effected by them, such as the character and composition of neighbourhoods and their communities (Baum-Snow and Marion, 2009).

One of these local effects concerns the composition of the housing stock, particularly the share of social housing within neighbourhoods. This share may shift due to targeted policies – such as urban renewal or mixed-tenure development – or through broader processes like selective migration and market dynamics. Changes in tenure composition can signal broader shifts in neighbourhood character. Residents may interpret such changes as early signs of shifts in social composition, reputation or safety, which can affect how desirable the neighbourhood is perceived to be (Permentier et al., 2009; Clark and Van Ham, 2009). As perceptions adjust, demand for housing may decline, putting downward pressure on prices. If enough buyers and sellers respond to these changing perceptions, this process may reinforce itself, leading to more pronounced price declines once a critical threshold is crossed, a dynamic commonly described as a tipping point (Schelling, 1971; Card et al., 2008; Malone, 2020). While theoretical models suggest that such tipping behaviour may occur, it remains an open empirical question whether such non-linear patterns can be observed in actual housing markets.

Neighbourhood attractiveness is more than a matter of individual taste or visual appeal. It affects where people choose to live, how long they stay and ultimately contributes to their quality of life (Sirgy and Cornwell, 2002). These perceptions are closely linked to outcomes such as health, education and access to opportunities (De Vries et al., 2003; Galster and Killen, 1995). Because neighbourhood composition shapes these perceptions, even modest changes in social housing share may alter how residents and buyers evaluate a neighbourhood's desirability. Understanding how differences in tenure structure influence these dynamics can therefore shed light not only on residential choices and neighbourhood change, but also on broader urban challenges such as segregation, inequality and spatial mismatch.

Since neighbourhood desirability cannot be observed directly, this study uses house prices as a proxy. As Rosen (1974) demonstrated, house prices reflect not only the attributes of the dwelling itself, but also those of the surrounding environment – including location, accessibility, amenities and social composition. These place-based attributes are often difficult to measure directly but are implicitly capitalised into what buyers are willing to pay. More recent studies confirm that house prices consistently incorporate preferences for local characteristics (Diamond and McQuade, 2019; Bell et al., 2023), making them a suitable market-based indicator of perceived neighbourhood attractiveness.

This study focuses on whether changes in the composition of the housing stock, particularly in the share of social housing, are associated with non-linear effects on neighbourhood desirability, proxied by house prices. In some urban processes, small shifts in composition can trigger disproportionate responses, often referred to as a tipping point (Card et al., 2008). In the context of this thesis, a tipping point means that as the share of social housing increases beyond a certain threshold, house prices in the corresponding neighbourhood decline more steeply. These kinds of dynamics are relevant because they suggest that the impact of housing decisions may not be evenly distributed and that certain thresholds matter more than others. This also carries practical implications: a tipping point would suggest that the local impact of new social housing depends not just on the absolute number of units, but on how they shift the perceived composition of a neighbourhood. To investigate whether the relationship between social housing share and neighbourhood desirability is linear, nonlinear or marked by a tipping point, this thesis is guided by the following main question:

Is the relationship between neighbourhood social housing share and local house prices in Amsterdam linear, gradually non-linear, or marked by a tipping point?

The aim is not to estimate causal effects in this relationship, but to investigate the shape of the association between the two and whether this shape indicates linearity, non-linearity or even the presence of a tipping point. Non-linearity in this context refers to deviations from a constant marginal effect, whether through a distinct slope change (a tipping point) or more gradual variation in the relationship.

In this study, three possible hypotheses are used to interpret the shape of the relationship between social housing share and neighbourhood house prices:

- $\mathbf{H_0}$: The relationship is linear and constant.
- $\mathbf{H_1}$: The relationship is non-linear, with a tipping point.
- $\mathbf{H_2}$: The relationship is non-linear but gradual, without a tipping point.

To explore these hypotheses, the study uses non-parametric and piecewise regression techniques applied to housing transaction data from Amsterdam, combined with neighbourhood-level indicators of social housing. To guide the analysis, the study uses three subquestions that each correspond to a key step in the empirical strategy. These subquestions are not answered separately or treated as stand-alone components. Rather, they serve as analytical building blocks that structure the empirical chapters and help interpret the results in relation to the main question.

1. What is the overall shape of the relationship between social housing share and house prices across Amsterdam's neighbourhoods?
2. Does the estimated relationship suggest non-linear effects or a tipping point in how social housing share relates to local house prices?
3. How robust is the estimated relationship to alternative model specifications, including fixed effects, between/within decomposition and lagged effects?

The empirical strategy is structured around these three subquestions. The first is addressed through exploratory visual methods, including LOESS smoothing and binned scatterplots, which offer an initial impression of the overall shape of the relationship without imposing a predefined form. To examine the second subquestion, the analysis turns to flexible regression models: Generalised Additive Models (GAMs) are used to detect gradual nonlinearities, while piecewise linear models test for the presence of slope changes consistent with a tipping point. Finally, the third subquestion is explored through a series of robustness checks.

At its core, this thesis is about understanding what makes neighbourhoods attractive places to live. Neighbourhood desirability is not just a matter of taste. It shapes where people want to live, how long they stay and how they experience daily life (Permentier et al., 2009). From a societal perspective, this matters because attractive and liveable neighbourhoods contribute to health, well-being and social cohesion (Sirgy and Cornwell 2002; De Vries et al. 2003). If we want to improve the quality of urban environments, we need to understand how different housing compositions influence how neighbourhoods are perceived. Investigating the role of social housing in this process helps clarify how planning and policy decisions affect not only affordability, but also the broader lived experience of city life. This is especially relevant in cities like Amsterdam, where spatial pressures and housing needs closely interact.

The findings of this study also carry direct implications for policy. If the relationship between social housing share and neighbourhood desirability is linear and constant (H_0), it suggests that every additional unit has the same marginal effect across all neighbourhoods. In that case, even distribution could be a straightforward goal. If there is a tipping point (H_1), then exceeding a certain threshold might disproportionately affect neighbourhood appeal, meaning that policymakers may wish to avoid crossing this threshold. If the effect is gradual but non-linear (H_2), then outcomes depend on local context, and a flexible and design-sensitive approach is more appropriate than applying uniform targets everywhere.

The remainder of this thesis is structured as follows. Chapter 2 reviews the literature on social housing, neighbourhood change and non-linear effects. Chapter 3 and 4 describe the data and methods respectively. Chapter 5 presents the empirical results. Chapter 6 discusses these findings in light of the existing literature, reflects on limitations and suggests directions for further research. Chapter 7 concludes.

2. Theoretical framework

Chapter two reviews the academic literature that informs the central questions of this thesis. It brings together insights from different strands of work in housing economics and urban studies, with the aim of situating the research in a broader theoretical and empirical context. The chapter closes by identifying a set of gaps in the literature that motivate the empirical approach taken in the rest of the thesis.

2.1 Literature review

This first section discusses existing academic work relevant to this study. It covers three main areas: the interaction of social housing, house prices and neighbourhood dynamics, the idea of tipping points and other non-linear dynamics in neighbourhood change and relevant estimation strategies.

2.1.1 Social housing, house prices and neighbourhood dynamics

Social housing plays a central role in shaping the spatial and social structure of cities. While its primary aim is to provide affordable housing for lower-income households, the presence and distribution of social housing also influence broader neighbourhood dynamics, including segregation patterns, local investment and residential mobility (Nguyen, 2005; Diamond and McQuade, 2019). These wider effects have made social housing an important object of study not only in housing policy, but also in urban economics and spatial planning.

Empirical studies show that the local effects of social or affordable housing developments vary strongly by neighbourhood context. Nguyen (2005), in a review of U.S. based studies, highlights that these effects depend heavily on design, scale and local circumstances. Poorly managed projects, visible deterioration or a mismatch with surrounding land uses are often associated with negative price effects. In several U.S. cases, price declines of between -2% and -10% have been observed near such developments. In contrast, smaller, well-designed projects that integrate smoothly into the existing neighbourhood fabric tend to have neutral or even positive effects, with price changes typically ranging from 0% to +3%. Nguyen further notes that early studies reporting large negative impacts often focused on extreme or highly visible problem cases, while more recent research finds more nuanced outcomes depending on local conditions.

Diamond and McQuade (2019) provide large-scale evidence on the spatial spillovers of housing developments funded by the Low-Income Housing Tax Credit (LIHTC) programme across U.S. cities. They find that in low-income neighbourhoods, the introduction of affordable housing leads to measurable improvements: house prices increase by around 6.5%, crime rates fall and the local population becomes more economically and racially mixed. In contrast, in high-income neighbourhoods, new LIHTC projects are associated with small price decreases of approximately 2% to 4%, alongside a shift toward lower-income residents. These findings suggest that affordable housing does not have a uniform effect but rather interacts with neighbourhood composition to shape housing markets and residential mobility.

While much of the empirical literature on neighbourhood effects of affordable housing focuses on the U.S., similar issues have also been discussed in the Dutch context. In Dutch cities, the role of social housing has been shaped by a different institutional setup, where housing associations have traditionally provided a large share of the total stock and served a broader part of the population (Van Gent and Hochstenbach, 2019). A recent study from Buurma-Olsen et al. (2025) shows that this system has come under increasing strain, with rising demand, stricter allocation rules and a growing mismatch between the social housing stock and the needs of eligible households. They estimate that around 30% of social housing units are currently occupied by households that exceed the official income eligibility thresholds, while waiting lists for eligible households have lengthened sharply in high-demand cities. These shifts have made it more likely that low-income households end up concentrated in certain areas. What this means for local housing markets is still unclear. Despite growing debate, there's little empirical work that looks at the relationship between social housing share and house prices in Dutch neighbourhoods.

Part of this gap may have to do with how the role of social housing has changed over time. Boelhouwer (2018) argues that recent reforms – like the 2015 Dutch Housing Act – have narrowed the function of the social housing sector in the Netherlands, shifting it from a general provision model toward a more residual one focused on lower-income groups. According to the paper, this shift has not only changed who can access social housing but also made it more likely that lower-income households become concentrated in specific areas.

Favilukis et al. (2023) approach the question from a different angle. Using a general equilibrium model calibrated to New York City, they simulate how different affordable housing policies affect not only overall supply but also the spatial distribution of households and neighbourhood composition. Their simulations suggest that depending on the design of the subsidy scheme, up to 10% of low-income households may relocate across neighbourhoods. At the same time, price effects vary across the city, with some neighbourhoods experiencing price declines of up to 3% following affordable housing interventions. These results illustrate how housing policy can reshape not just affordability, but also the broader spatial allocation of income groups across urban areas.

Similar context-dependent variation is found in regeneration settings. A recent study by Neumann and Yasar (2024) shows how this variation plays out in the context of a large urban regeneration programme in Germany. They find that the redevelopment of social housing in disadvantaged areas led to rising nearby property values, but only when combined with broader improvements in infrastructure and public space. In these cases, house prices increased by approximately 4% to 8%. In contrast, projects that focused narrowly on housing renewal without addressing the wider urban environment had little to no price effect. This suggests that the impact of social housing on house prices cannot be separated from the surrounding neighbourhood context and accompanying policy measures. Another similar point is made by Van Gent and Hochstenbach (2019), who argue that in Dutch cities, the spatial distribution of social housing interacts with gentrification processes and housing market segmentation, with property values in gentrifying Amsterdam neighbourhoods rising by 20% to 30% over the past decade as social housing shares declined.

While these studies provide useful insights, most focus on the effects of new affordable or social housing developments, often in the context of regeneration programmes or policy interventions. However, the empirical research of this thesis focuses on how the existing distribution of social housing across neighbourhoods relates to local house prices – an area that has received far less empirical attention. In particular, this paper examines whether the relationship between neighbourhood social housing share and local house prices is linear or whether there are threshold effects, where the impact of social housing becomes more pronounced beyond a certain share. This idea connects to a broader literature on tipping points in neighbourhood change, which suggests that small shifts in composition can trigger larger changes in outcomes once a critical threshold is crossed (Card et al., 2008; Malone, 2020).

2.1.2 *Tipping points and non-linear neighbourhood dynamics*

The concept of neighbourhood tipping points originates from Schelling's (1971) early work on segregation dynamics, which showed how individual residential preferences can lead to large-scale, self-reinforcing shifts in neighbourhood composition. Using a simple grid-based simulation, Schelling demonstrated that even mild preferences for same-group neighbours could lead to highly segregated outcomes. In his model, once the share of minority residents in a neighbourhood passed a certain point, majority residents would start to leave, triggering a chain reaction that ultimately reshaped the entire neighbourhood.

Building on Schelling's idea, later research has shown how small shifts in neighbourhood makeup can sometimes lead to much bigger changes. Card et al. (2008) find that in U.S. cities, white residents often start to move out once the minority population in a neighbourhood passes a certain point, usually somewhere between 5 and 20 percent. This kind of "white flight" supports the idea that people may tolerate gradual change up to a point, but once a threshold is crossed, reactions become stronger and neighbourhoods can change quickly.

Bisin et al. (2011) help explain how this process works. They build a model showing that even when people have fairly smooth or moderate preferences, sharp shifts can still happen. What matters is not just who lives in a neighbourhood, but how people expect others to behave. If people think others will leave, they may do the same. These expectations can lead to tipping, even when no one has a strict personal cut-off. In other words, large changes can grow out of many small, cautious decisions.

We see this pattern in other contexts as well. Saiz and Wachter (2011) look at immigrant inflows and find that native-born households do not respond much to small increases but begin to leave more often once the share of immigrants reaches a higher level. The pattern is not steady or proportional. Instead, change speeds up after a certain point. This supports the idea that neighbourhood composition can remain stable for a while, then shift quickly once enough residents start reacting.

A more recent study has shown that tipping behaviour is not limited to race or ethnicity. Malone (2020) looks at income-based tipping in U.S. metropolitan areas. He finds that when the share of very low-income households in a neighbourhood rises above roughly 10%, higher-income residents are more likely to leave. This pattern is especially visible in areas with many families and school-age children, which suggests that concerns about local amenities or school quality may play a role. Malone's findings suggest that perceived declines in neighbourhood quality – even in the absence of racial change – can activate similar out-migration dynamics.

While most tipping studies focus on the U.S., some work has looked at whether similar patterns emerge in European contexts. Ong (2017) investigates whether tipping points in ethnic composition exist in Dutch urban neighbourhoods and finds that while some threshold effects are observable, they tend to be more gradual and context-dependent than in U.S. studies. This may reflect differences in institutional context, although Ong does not identify specific mechanisms. Other European studies, such as Zorlu and Latten (2009), and Andersson and Bråmå (2018), describe selective out-migration by native residents in response to rising shares of immigrants or low-income groups. While these dynamics are not framed explicitly as tipping points, they show that similar sorting mechanisms may operate in European settings (Dutch and Swedish respectively), albeit more gradually.

However, tipping dynamics are not an inevitable outcome of residential preferences. Bruch and Mare (2006) use simulation-based modelling to show that tipping points only emerge under certain assumptions, such as strong conformity pressures or sharp threshold-based preferences. In many plausible settings, neighbourhood change takes place more gradually, with demographic shifts occurring smoothly even as preferences vary. Their findings underscore the importance of carefully testing for non-linearities, rather than assuming sharp breaks beforehand.

While tipping points are not guaranteed to occur, it is still useful to define what qualifies as one in empirical terms when studying them. In this thesis, a tipping point refers to a discrete threshold in the neighbourhood share of social housing, beyond which the relationship with local house prices becomes, both statistically and economically, significantly stronger or weaker. Although previous studies often anticipate a negative association, a tipping point could in principle mark a sharp change in either direction. This definition follows the approach of Card et al. (2008) and Malone (2020), who identify tipping behaviour through structural breaks in slope rather than assuming specific effects beforehand. A point is considered a tipping point only if the change in slope is both statistically significant and substantively large and should suggest an abrupt shift rather than a gradual trend. This distinction helps separate genuine tipping dynamics from smoother non-linear patterns.

Taken together, the studies in this section suggest that neighbourhood change can be non-linear, shaped by feedback effects that emerge once key thresholds in composition or perception are crossed. While most empirical work so far has focused on population characteristics such as ethnicity or income, less is known about the presence of tipping points in housing markets.

2.1.3 Tipping points in house prices

In housing research, the concept of a tipping point often refers to a critical threshold in the presence of a particular group or housing type, such as social or subsidised housing, beyond which neighbourhood dynamics change in a non-linear way. Although most papers on tipping points focus on demographic sorting or changes in racial composition, several studies have explored whether house prices themselves respond with a clear break or shift around such thresholds.

Card et al. (2008), in their influential analysis of white flight in American cities, examined whether neighbourhoods passed a racial tipping point that triggered rapid out-migration and falling house prices. They found strong evidence of tipping in demographic composition, but only modest effects on property values. Their results suggest that, although people may respond abruptly to changes in neighbourhood composition, house prices tend to adjust more gradually. This indicates that market-based reactions may be less sensitive or less immediate than behavioural ones.

Other studies do suggest that house prices can respond in a more threshold-like way, particularly when subsidised housing becomes spatially concentrated. For example, Galster et al. (2001) analysed small clusters of subsidised rental homes spread throughout different residential areas in Denver. They found that having a small number of subsidised units nearby, up to about five within a 500-foot radius, was associated with slightly higher property values. But once that number rose above six, home prices in the area began to decline. The authors interpret this as evidence that a few well-managed units may have little effect, but that concentration can tip neighbourhood perception and trigger a decline in demand. A similar pattern was observed by Lyons and Loveridge (1993) in Minnesota. While the addition of a single subsidised unit had a minimal effect on nearby property values, placing an entire low-income housing project in the same area was associated with an average price drop of over 1,600 dollars. Although they did not explicitly test for a threshold effect, their findings suggest that the scale and visibility of subsidised housing may influence the magnitude of nearby price responses.

Outside the United States, there are fewer studies, but some cases show a similar pattern. In South Africa, Du Preez and Sale (2011) studied the effect of a new social housing development in Port Elizabeth. They found that the introduction of a large low-cost housing complex significantly reduced nearby property values. Although the study did not pinpoint a specific tipping point, it supports the broader concern that highly visible concentrations of low-income housing can lead to negative reactions in the housing market.

Together, these findings suggest that the effects of social or affordable housing on property values may depend not just on the number of units, but on their spatial concentration and visibility. A small number of dispersed units may have little or even a positive effect, while larger concentrations can trigger negative market responses. This makes it important to consider whether and when tipping points occur. To investigate such patterns, researchers increasingly rely on flexible estimation approaches that can detect non-linearities or structural breaks in the relationship between housing attributes and prices.

2.1.4 Flexible and threshold-based estimation approaches

Most empirical studies of house prices start from the hedonic pricing model, which assumes that a property's value reflects the characteristics of the dwelling and its location. The underlying framework was introduced by Rosen (1974), who showed how observed market prices can be used to estimate the value of individual product attributes in heterogeneous markets. In this setting, housing is not seen as a uniform good, but as a bundle of characteristics – such as floor area, number of rooms, age, location, architectural design – each of which contributes to the overall price. By observing how prices vary across properties with different attributes, researchers can infer how much buyers are willing to pay for each characteristic. This approach has been widely adopted in urban economics and housing market research, where it forms the basis for many regression-based models of price variation.

In this context, house prices are often used as a proxy for neighbourhood desirability. As Rosen (1974) explains, prices reflect not only the features of individual dwellings, but also the characteristics of their surroundings, such as location, accessibility, amenities, and social composition. These factors are difficult to observe directly but are implicitly priced into transactions. This makes house prices a useful summary measure of how desirable an area is in the eyes of buyers. While they do not capture every aspect of neighbourhood quality, they provide a consistent market-based signal of relative demand across space, which is why they are commonly used in studies of urban dynamics and neighbourhood change (Bell et al., 2023).

Sirmans et al. (2005) provide an extensive review of how the hedonic pricing model has been applied in housing market research. They show that, while the core structure remains consistent, the choice of variables, functional form and modelling techniques can vary widely depending on the context and available data. For example, they find that studies on U.S. data often include school district quality on crime rates, while research in other contexts may focus more on building age or tenure type. For functional form, in areas with high price variation, semi-log or non-linear functional forms are more common, since they are better at capturing diminishing marginal returns to housing attributes like floor space.

These choices are not just technical decisions. They depend on how housing markets work in practice. In cities with a mix of housing types, policy influences or big differences between neighbourhoods, a simple linear model might miss important patterns. More flexible approaches make it easier to pick up non-linear effects, like shifts in the relationship once a variable passes a certain point. This is especially useful in a city like Amsterdam, where the share of social housing varies widely between areas, and where market conditions can differ quite a lot from one neighbourhood to the next.

To get a sense of how such variation plays out in the data, a useful starting point is LOESS (Locally Estimated Scatterplot Smoothing): a non-parametric method that fits smooth curves through noisy relationships without assuming a specific functional form. This approach was first introduced by Cleveland (1979), but it is now mostly used as an exploratory tool in applied settings. As Wood (2017) explains, LOESS works by fitting simple models to small, overlapping parts of the data, which gives more weight to observations that are closer to the point being estimated. This makes it useful for spotting broad patterns or non-linear relationships without imposing a fixed structure.

Another simple but effective way to explore patterns in the data is the binned scatterplot. Instead of plotting all raw observations, the data are grouped into equally sized bins based on the independent variable and the average value of the dependent variable is plotted for each bin. This reduces noise and makes it easier to see general trends, especially in large datasets. Diamond and McQuade (2019), for example, use this method to visualise how house prices vary with distance from affordable housing developments. This helped them detect non-linear effects before specifying a formal model. While binned plots are less flexible than LOESS curves, they offer a transparent way to visualise the average relationship across the distribution, and they help identify parts of the data where observations are sparse or unevenly distributed.

Binned plots and LOESS curves are useful for spotting general patterns, but they do not account for other covariates or fixed effects. A more formal way to estimate smooth, non-linear relationships is the Generalized Additive Model (GAM). As Wood (2017) explains, GAMs extend the linear regression framework by allowing certain predictors to enter the model as smooth functions rather than fixed coefficients. This provides more flexibility in capturing unknown or irregular patterns in the data, especially when the true relationship is unlikely to be strictly linear. GAMs are commonly used in applied work where the shape of the relationship is of interest but should be estimated from the data rather than imposed in advance.

Another way to model non-linear relationships is by using a piecewise linear model (PLM), which allows for a change in slope at one or more breakpoints. Rather than estimating a smooth curve, PLMs divide the range of a variable into segments and fit separate linear relationships within each segment. Card et al. (2008) use this approach to study tipping points in U.S. neighbourhoods, showing how white population decline accelerates once the minority share passes a certain threshold. Unlike GAMs, which let the data determine the overall shape of the curve, PLMs impose a simpler structure with clear breakpoints and slope changes. This makes them easier to interpret, especially when the goal is to test whether a threshold effect exists. In practice, both models can be used alongside each other. GAMs can be used to explore the shape of the relationship flexibly, while PLMs offer a way to test for specific threshold effects in a more structured way. This combination is especially useful in settings where theory suggests a potential tipping point, but the location and nature of that point need to be inferred from the data.

While flexible and segmented models help uncover non-linear relationships, they do not address the possibility that unobserved neighbourhood characteristics may bias the results. One common solution is to include neighbourhood fixed effects. These control for time-invariant differences across areas that might affect both housing prices and social housing share, such as school quality, historical development patterns or local institutions. When combined with time fixed effects, this creates a two-way fixed effects (TWFE) specification that controls for both neighbourhood-level heterogeneity and broader temporal trends. However, this approach comes with a trade-off: it eliminates between-neighbourhood variation, which is often informative in studies of spatial inequality.

To address this, researchers sometimes use a Mundlak-style model, which adds the group-level mean of time-varying covariates as an additional regressor. This method, discussed by Wooldridge (2021) among others, makes it possible to separate the effect of changes within neighbourhoods from differences between them, while still controlling for unobserved, time-invariant heterogeneity. In the context of housing markets, this helps distinguish whether price effects are driven by shifts over time within neighbourhoods or by longer-term structural differences across areas.

Lagged models provide a different kind of robustness check by testing whether the relationship between two variables takes place with a delay. For instance, the effect of a changing housing mix may not be immediate but may take several years to show up in transaction prices. Lagged strategies are commonly used in urban and housing research, where market responses are often gradual rather than instantaneous. For example, Autor et al. (2014) examine the effects of foreclosures on nearby property values in U.S. cities and include lagged foreclosure rates as part of their identification strategy to account for the delayed price effects of housing distress. Their findings show that the impact of foreclosures on house prices unfolds gradually over time and underline the importance of accounting for lag structure in urban housing market analysis.

Another way to isolate the effect of neighbourhood context is through a hedonic residual model, which removes variation in prices explained by housing structure before estimating the role of local composition. This is done in two steps: first, transaction prices are regressed on physical dwelling characteristics and second, the resulting residuals are analysed as a function of neighbourhood variables such as social housing share. The idea is that, by stripping out variation due to size, type and age, the remaining price differences reflect location-specific factors that may otherwise be confounded. This method has been used in housing research to focus more clearly on environmental or social characteristics that are difficult to measure directly (Gibbons and Machin, 2008). While it does not solve endogeneity, it helps clarify which patterns in the data are likely attributable to housing structure and which might stem from the local context.

Finally, local polynomial regressions provide a sharp, non-parametric check for discontinuities in the relationship. Instead of estimating the entire curve, this method focuses on whether there is a statistically significant jump at a specific threshold, such as a suspected tipping point. By fitting separate regressions on either side of a cutoff and comparing the intercepts, it tests for abrupt changes in the dependent variable without imposing a global structure on the data. This approach is closely related to regression discontinuity designs, although it does not require a running variable in the strict causal sense (Imbens and Lemieux, 2008). In the context of neighbourhood composition, it can be used to verify whether the relationship changes sharply at points identified by other models.

Together, these approaches reflect a growing recognition in urban economics that housing market relationships are often non-linear, non-constant or non-immediate. Whether through flexible modelling, threshold detection or robustness checks that account for unobserved context and timing, recent work increasingly shows the need to match empirical methods to the complexity of urban data.

2.2 Research gap

While much has been written about the role of social housing in shaping neighbourhood conditions, most of the empirical work has focused on new developments or regeneration programmes. Studies like those by Diamond and McQuade (2019) and Nguyen (2005) explore how the construction or renovation of affordable housing affects nearby property values, social mix or crime rates. These contributions offer valuable insights, but they often centre on short-term interventions rather than the broader, structural presence of social housing in existing urban neighbourhoods. In European contexts, such as the Netherlands, where social housing makes up a significant share of the housing stock, the question is not just what happens when affordable housing is built – but how its long-term distribution shapes local housing markets.

At the same time, the literature on tipping points and neighbourhood change has mostly looked at demographic shifts, such as racial or income-based segregation (Card et al. 2008; Malone 2020; Saiz and Wachter 2011). These studies show that neighbourhood dynamics can be highly non-linear, with small compositional changes sometimes triggering larger behavioural responses. But so far, little attention has been paid to whether similar patterns exist in relation to housing market outcomes, like prices. Also, many of these studies focus on U.S. cities, where institutional settings differ substantially from European housing systems. Dutch cities, for example, combine a large social housing sector with strong planning institutions, which may affect both the likelihood and shape of tipping behaviour (Ong 2017).

Although existing work has drawn attention to the importance of spatial context, non-linearity and robust identification strategies, few studies have applied this full range of tools to the question of how social housing share relates to local house prices. Flexible models like GAMs, PLMs and Mundlak approaches are increasingly used to uncover non-linear or delayed effects. Yet their use in social housing studies remains scarce, particularly with European datasets over longer time periods.

This leaves an important gap in understanding how social housing relates to housing market outcomes beyond specific developments or short-term policy changes. By connecting insights from housing policy, neighbourhood dynamics and tipping point theory, this study aims to offer a more detailed picture of how the presence of social housing affects local house prices across the distribution, rather than just on average.

3. Data

This chapter describes the data that have been used in this research. Specifically, Subchapter 3.1 introduces the data sources and provides the rationale behind the selection of these particular sources, including their relevance to the research question and their suitability for the applied empirical strategy. Subchapter 3.2 describes the transformation process and the construction of the final analytical dataset. Finally, section 3.3 discusses any limitations of the selected data sources.

3.1 Data description

This study relies on two primary data sources that together enable a detailed analysis of how the share of social housing at the neighbourhood level relates to local house prices in Amsterdam.

The first dataset contains transaction data from the Dutch Association of Real Estate Agents (NVM), covering the period 2000 to 2023. Access to this dataset was granted through the university under a formal data usage agreement and was limited to transactions within the ten largest Dutch municipalities, including Amsterdam. The dataset includes detailed micro-level information on approximately 75% of all Dutch individual housing transactions, such as sales prices, structural characteristics (e.g. dwelling type, floor area, construction period), and precise location data (including addresses and multiple neighbourhood codes).

The second dataset consists of annual figures on the share of social housing units managed by housing associations (*corporatiewoningen* in Dutch) per neighbourhood (*buurt)* in Amsterdam from 2011 to 2024. These data were manually extracted from the *Dashboard Kerncijfers*, published by the Municipality of Amsterdam. The key explanatory variable in this study – share of social housing – was derived from this dataset.

Combined, these two datasets offer a high-resolution view of Amsterdam’s housing market, allowing for the analysis of neighbourhood-level social housing concentrations and their relationship to private housing prices over time.

3.2 Data transformation and construction

This section describes the main steps taken to transform the raw data into a structured, analysis-ready dataset. It outlines the manual construction of the social housing dataset, the cleaning and filtering of housing transaction data, and the procedure used to merge the two sources. All steps were carefully documented to ensure transparency and replicability, and can be found in the supporting materials.

3.2.1 Municipal data on social housing share

The second dataset, which contains annual neighbourhood-level shares of social housing in Amsterdam, was manually constructed using publicly available data from the *Dashboard Kerncijfers* published by the Municipality of Amsterdam. For each year from 2011 to 2024, relevant observations were retrieved by navigating the dashboard and selecting the appropriate theme, city-department, GGW-area, district, and neighbourhood. The value of interest – the percentage of *corporatiewoningen* (i.e. social housing units managed by housing associations) – was then recorded manually into an Excel file. This process was repeated for all relevant neighbourhood-year combinations. A document describing the exact extraction steps, including screenshots and selection instructions, is included as Supportive file 2.

3.2.2 Data cleaning and merging

The raw housing transaction data (NVM) was filtered to include only transactions within the municipality of Amsterdam. Since the dataset contains multiple variables referencing different geographical units over time, neighbourhood identity was standardised by using the 2022 CBS neighbourhood codes (buurt_nr_2022) throughout the analysis. This choice ensures consistency in spatial definitions across years and avoids issues caused by changing or reclassified neighbourhood boundaries. The filtered dataset is included as Dataset 3.

The social housing dataset was manually extracted from the Municipality of Amsterdam's dashboard, found on their website. Neighbourhood names were matched to CBS *buurtcodes* using an official mapping table (Dataset 4). Since the dashboard used local neighbourhood naming conventions, substantial manual cleaning was required. A mapping vector was constructed to recode incorrect or misspelled names. Five neighbourhoods that did not exist at the time of the CBS code registration (2022) were excluded from the dataset. The complete correction procedure is documented in R-script 1.

The two datasets were then merged on CBS buurtcode and year. Since neighbourhood-level data on social housing share is only available from 2011 onwards, the regression sample is limited to the period 2011–2023. Transactions outside this range – those from before 2011 or after 2023 – were kept for exploratory purposes but excluded from the main analysis due to missing values for the key explanatory variable (2000 – 2010) or the key outcome variable (2024). This ensures that all included observations are based on consistent and complete data.

Prices below €100,000 were excluded because they fall below the first percentile and likely reflect non-standard transactions, such as symbolic sales or data errors. Prices above €2,100,000 were also removed, as they exceed the 99th percentile and represent the extreme high end of the housing market, which could distort the analysis. A natural logarithm of the transaction price was created and stored as a new variable (log_price) to reduce skewness in the price distribution and allow for a more stable regression specification. The main explanatory variable – the share of social housing in a neighbourhood – was rescaled from a percentage to a proportion between 0 and 1 to better capture potential non-linear effects and improve interpretability.

To avoid distortion from sparsely populated extremes, the analysis excludes transactions in neighbourhoods where the share of social housing exceeds 90%. This affects only 0.65% of observations but removes extreme cases with limited data support. All main regression models are estimated on the trimmed sample.

To improve clarity and readability throughout the analysis, the original Dutch variable names were renamed to consistent English equivalents. For instance, *gebruiksoppervlaktewoonfunctie* became *living_area_m2*, and *woningtype* was renamed to *dwelling_type*. This step was applied across the entire cleaned and merged dataset.

To further ensure internal consistency, the dataset was restricted to transactions with complete information on a selected set of control variables. These controls were chosen based on both their economic relevance and their data availability. Several structural housing variables were available in the raw dataset but were excluded from the final model due to high rates of missing values or inconsistent measurement. For example, housing subtype and plot size were missing in approximately 88% of observations, while variables such as external storage area and other internal space had missingness rates of around 50% and 84%, respectively. Although gross volume was largely complete, it was found to be highly collinear with living area and added limited explanatory power. Including these variables would have led to a substantial reduction in sample size and potentially biased subsample estimates.

The final specification includes living area (*living_area_m2*), dwelling type (*dwelling_type*), number of rooms (*num_rooms*), and construction period (*construction_period*). These variables had either no or minimal missingness and capture key structural features of dwellings that influence house prices. The variable *construction_period* was included as a categorical indicator of building age and type. However, the underlying codes showed considerable overlap in construction years and lacked clear documentation. As a result, the categories are treated as fuzzy source groupings, not interpreted as chronological periods. Leaving this variable out of the analysis was considered too, but a model comparison showed that including this variable improves explanatory power.

The resulting dataset includes 104,716 transactions and forms the empirical basis for the analysis in the following chapters. It is available as Dataset 6. All steps used for the cleaning and merging process are documented in R-script 1.

3.3 Data limitations

While the datasets used in this study offer rich spatial and temporal detail, several limitations should be acknowledged. First, this study uses neighbourhood-level data based on CBS definitions, which may not fully capture the more fine-grained scale at which homebuyers form perceptions. This introduces the risk of spatial mismatch, often referred to as the Modifiable Areal Unit Problem (Openshaw, 1984). Still, neighbourhoods offer a practical and policy-relevant level of analysis, and the use of over 500 distinct units within Amsterdam provides relatively detailed spatial resolution.

Second, the social housing dataset was constructed manually from an online dashboard and is limited to the period 2011 – 2024. No historical data are available for earlier years, restricting the scope of the merged dataset. Although housing transactions prior to 2011 are retained for exploratory use, they are excluded from the main regression analysis due to the lack of corresponding social housing figures.

Furthermore, the use of 2022 CBS neighbourhood codes for all years assumes that spatial boundaries remained stable over time. While this choice ensures consistency in neighbourhood definitions, it does not account for boundary changes, splits or reclassifications that may have occurred between 2011 and 2023. However, an inspection of the number of unique neighbourhood codes shows only a modest increase – from 375 in 2011 to 390 in 2023 – suggesting that spatial definitions remained largely stable during the period used for regression analysis.

Some potentially useful housing characteristics had to be left out due to too many missing values. This includes variables like housing subtype, plot size and gross volume, which were incomplete for large parts of the dataset. Including them would have led to a much smaller and less consistent sample. Still, there were enough relevant variables without missing values to account for key differences between dwellings.

Finally, the process of matching neighbourhood names to CBS codes involved several manual corrections and exclusions. Although these steps were performed systematically and documented thoroughly (see R-script 1), there remains a possibility of residual mismatches or imprecision in the geographic linkage.

3.4 Descriptive statistics

This section provides a brief descriptive overview of the dataset. It focuses on the dependent variable (log-transformed transaction price), the main explanatory variable (neighbourhood-level share of social housing), and the structural housing characteristics used as controls. The data described here have already been cleaned and merged, with outliers removed by trimming the lowest and highest percentiles of transaction prices. This overview helps clarify the scope of the data, highlight any notable patterns, and justify the modelling choices made in later sections.

The dependent variable in all models, except the two-stage hedonic specification, is the natural logarithm of transaction price. This transformation reduces the right-skewness typically found in raw house price data and allows regression coefficients to be interpreted approximately as percentage changes. The mean log price is 12.83, corresponding to roughly €376,000. The values range from 11.51 to 14.56, which translates to a minimum of about €100,000 and a maximum of around €2.1 million. As shown in Figure 3.1, the distribution is approximately symmetric and bell-shaped, supporting the use of log price in both linear and flexible regression models.

Figure 3.1: Distribution of log-transformed prices (2011 - 2023)

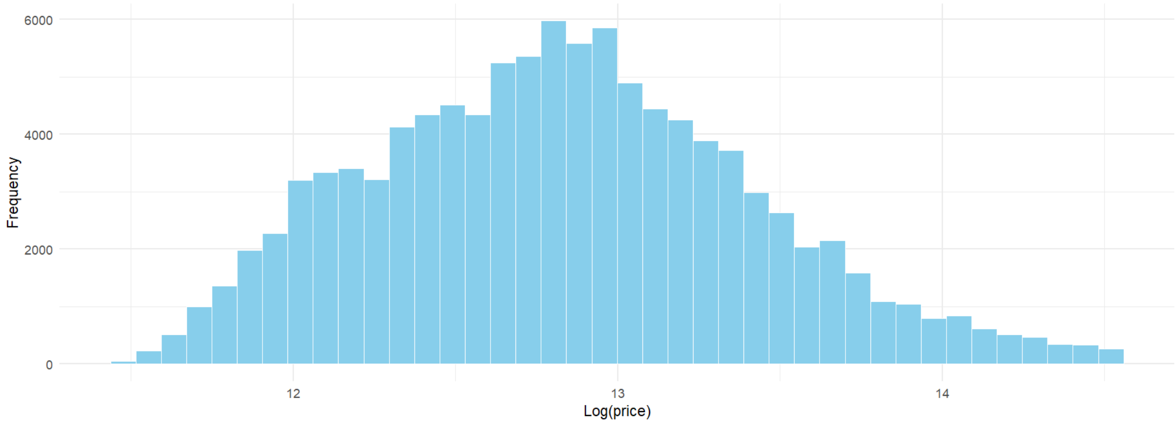
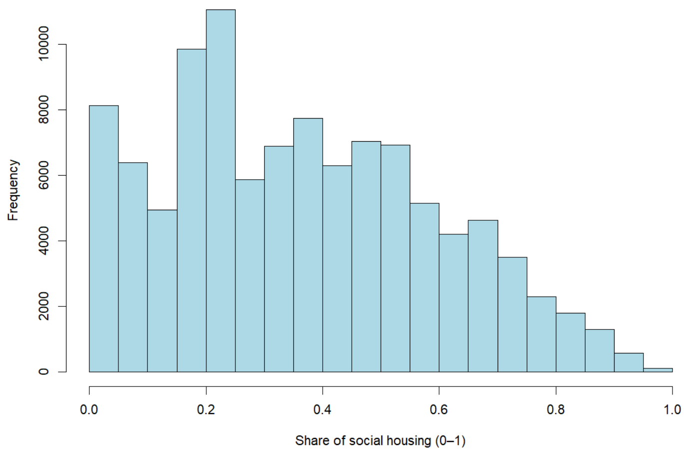


Figure 3.2 shows that the distribution of neighbourhood-level social housing share is skewed toward the lower and middle parts of the range. Most housing transactions occur in areas where the share falls between 0% and 70%, with a clear concentration in the lower half. Observations become sparse above 90 percent because only a small number of neighbourhoods have such high social housing shares and those neighbourhoods account for relatively few transactions.

Figure 3.2: Distribution of neighbourhoods by social housing share, showing a concentration in the lower and middle ranges and relatively few cases above 90%.



Because of this sparsity at the upper end of the distribution, all main regression models are estimated on a trimmed sample that excludes neighbourhoods with social housing shares above 90%. This ensures that estimated effects are based on a part of the distribution where the data are more reliable. For the purposes of visual exploration – including descriptive plots and smooth functions – the full sample is retained to give a complete picture of the data.

4. Methodology

This chapter describes the empirical strategy for analysing the relationship between the neighbourhood share of social housing and individual house prices in Amsterdam. The central aim is to test whether this relationship is linear or shows signs of a tipping point – a threshold beyond which the effect of social housing becomes more strongly negative.

The following hypotheses guide the analysis:

- **H_0**: The relationship is linear and constant.
- **H_1**: The relationship is non-linear, with a tipping point.
- **H_2**: The relationship is non-linear but gradual, without a tipping point.

To evaluate these hypotheses, the chapter introduces a flexible regression model that allows for non-linear effects of social housing share, while controlling for key structural housing characteristics and time fixed effects. First, the relationship is estimated using a set of simple visual tools to explore its general shape (4.1). Then, the baseline specification is described in section 4.2, followed by a discussion of control variables (4.3), estimation approach (4.4), (4.5) and identification strategy and assumptions (4.6).

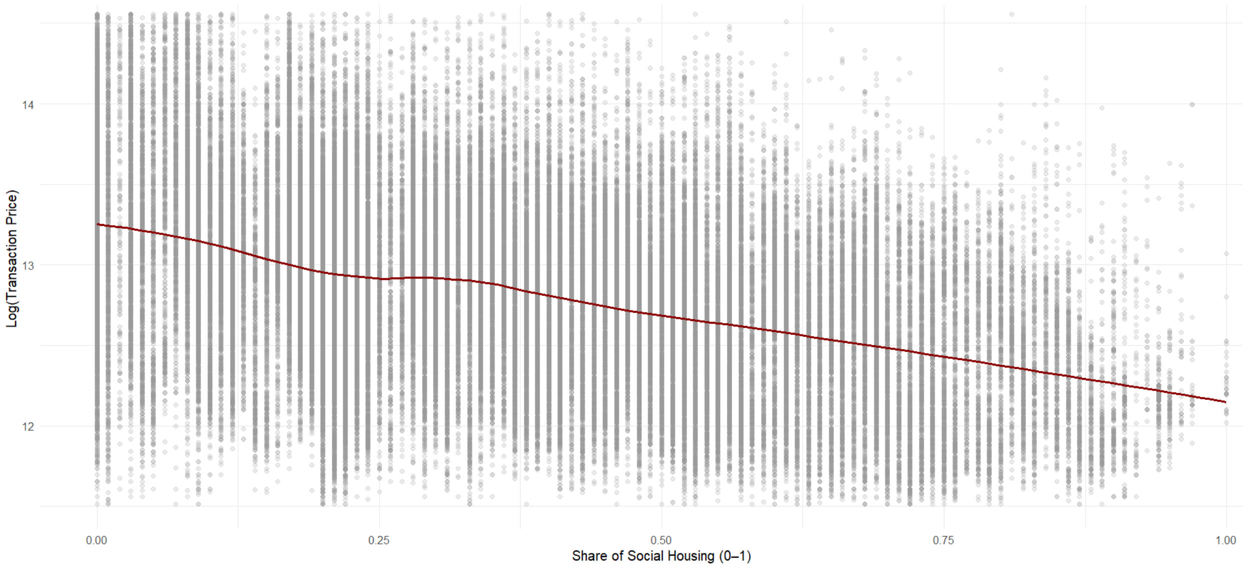
4.1 Visual exploration

Before describing the baseline models, the relationship between social housing share and house prices is first explored visually. This helps assess whether a linear specification is appropriate or whether a more flexible approach is needed.

4.1.1 LOESS

Figure 4.1 shows a Locally Estimated Scatterplot Smoothing (LOESS) curve for the relationship between the neighbourhood-level share of social housing (measured annually from 2011 to 2023) and the logarithm of housing prices across all transactions in Amsterdam. The span parameter was set to 0.5 to allow for moderate local sensitivity without overfitting the data. A version using the default, more smoothed LOESS curve is included as Appendix 1 for comparison.

Figure 4.1: LOESS-smoothed relationship between social housing share and log house prices. Span = 0.5



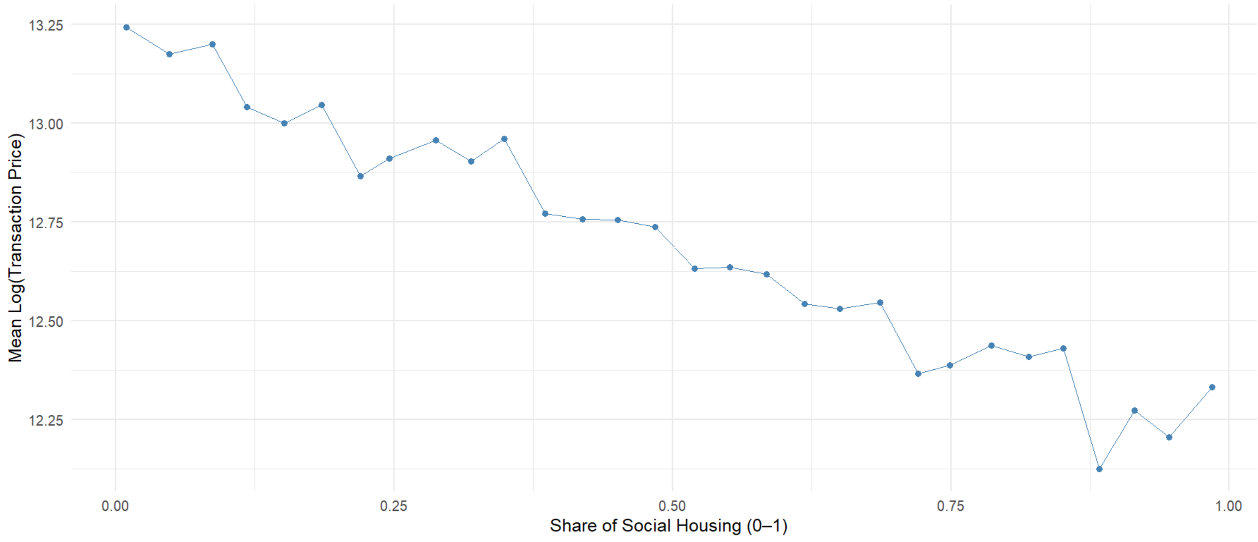
The trend is clearly negative across the full range: house prices are lower in areas with more social housing. The slope is relatively linear at higher shares of social housing but varies more for low shares. There is no clear indication that prices decrease more or less sharply at lower social housing shares; the relationship in that range appears more variable rather than systematically different.

The fluctuation at lower social housing shares may reflect greater variation in neighbourhood characteristics within this range, or differences in how small concentrations of social housing interact with local market outcomes. In contrast, the pattern becomes more stable and predictable as the share increases. This visual evidence supports the idea that a flexible specification, rather than a simple linear model, is appropriate for capturing the full shape of the relationship. The overall downward trend is remarkably consistent across years, as shown in Appendix 2, which displays the LOESS-smoothed relationship separately for each year from 2011 to 2023.

4.1.2 Binned scatterplot

Figure 4.2 shows a binned scatterplot of the relationship between social housing share and the logarithm of transaction prices. Each point represents the average log price within a bin of neighbourhood-level social housing share. The pattern confirms the negative relationship: average prices tend to be lower in neighbourhoods with more social housing.

Figure 4.2: Binned scatterplot: relationship between social housing share and log house prices. Bins = 30



The decline is most consistent between approximately 30% and 70% social housing, where the trend appears relatively smooth and downward sloping. At lower shares, the relationship fluctuates more, which aligns with the pattern seen in the LOESS plot. Toward the upper end of the distribution, the slope increases slightly. This may reflect noise due to fewer observations or due to idiosyncratic neighbourhoods with very high social housing concentrations.

The LOESS and binned scatterplot both indicate a non-linear pattern in the relationship between social housing share and house prices. This suggests that relying on a fixed linear specification could misrepresent the underlying association. A flexible estimation approach is therefore more appropriate for the main analysis in this study.

4.2 Baseline model specification

The analysis starts with a flexible regression model that captures how local house prices vary with the share of social housing in the neighbourhood. The goal is to allow for a non-linear relationship, without imposing a fixed shape in advance. This model is specified as:

$$\ln(P_{it}) = \alpha + f\ (S_{nt}) + \boldsymbol{X'}_{it}\beta + \ \delta_t + \ \varepsilon_{it}$$

Where:

$\ln(P_{it})$	= natural logarithm of the transaction price of dwelling i in year t
$\alpha =$	= intercept
$f\ (S_{nt})$	= smooth, flexible function of the share of social housing in neighbourhood n in year t
$\boldsymbol{X'}_{\mathrm{it}}\beta$	= vector of housing characteristics and their associated coefficients
δ_t	= year fixed effects
ε_{it}	= error term, unobserved factors

The vector of controls $\mathrm{X'}_{\mathrm{it}}$ includes living area (*living_area_m2*), dwelling type (*dwelling_type*), number of rooms (*num_rooms*), and construction period (*construction_period*). These control variables are further justified in Chapter 4.2.

The function $f\ (S_{nt})$ will be estimated non-parametrically to allow for a flexible and potentially non-linear relationship between social housing share and house prices. For the baseline model three alternative modelling strategies are applied:

1. A **Generalised Additive Model (GAM)** using penalised smoothing splines. This flexible model captures gradual non-linear patterns without requiring a pre-defined threshold.
2. A **Piecewise linear model** with a constrained algorithmic breakpoint search**,** which tests whether the slope changes at a particular point in the distribution. This provides a simpler, more interpretable test of tipping behaviour.
3. A **Piecewise linear model** with a manual grid search, which systematically estimates the model across a range of candidate breakpoints to transparently compare slope changes and model fit.

4.3 Control variables

To isolate the relationship between neighbourhood-level social housing share and house prices, the model includes a set of structural housing characteristics as control variables. These characteristics account for variation in housing quality and composition that directly influence transaction prices.

Living area (*living_area_m2*): Larger dwellings generally command higher prices, often with a near-logarithmic relationship (Sirmans et al., 2005). This variable serves as the most direct proxy for dwelling size and is central to hedonic price models.

Dwelling type (*dwelling_type*): Different dwelling types (e.g. apartments, terraced houses, detached homes) vary systematically in terms of size, layout, location and desirability (Sirmans et al., 2005). Including a categorical indicator for dwelling type helps account for these structural differences.

Number of rooms (*num_rooms*): The number of separate rooms in a dwelling affects perceived functionality and attractiveness, especially for families (Malpezzi, 2002). It also captures layout-related variation that is not explained by floor area alone.

Construction period (*construction_period*): This categorical variable groups dwellings by construction period, reflecting differences in architectural style, building quality, and energy performance (Malpezzi, 2002). Although the exact definitions of the categories are unclear and show considerable overlap in construction years, model comparison showed that including this variable improves explanatory power. As such, it is retained as a control to account for systematic variation in housing characteristics linked to building age.

All four variables are commonly used in housing literature to control for heterogeneity in structural quality (Sirmans et al., 2005; Malpezzi, 2002). In addition to their theoretical relevance, they were selected for their empirical completeness: missing values were either zero or minimal (29 missing values in one case, the rest 0 missing values) across the full sample, which allows for a large and consistent dataset.

4.4 Estimation strategy

To explore the shape of the relationship, both flexible and segmented estimation methods are applied, allowing the data to inform whether any sharp structural break exists.

4.4.1 Generalised Additive Model (GAM)

To allow for a flexible, data-driven relationship between the neighbourhood share of social housing and house prices, the baseline model is estimated using a Generalised Additive Model (GAM). The specification is as follows:

$$\ln(P_{it}) = \alpha + f\ (S_{nt}) + \boldsymbol{X'}_{\mathrm{it}}\beta + \ \delta_t + \ \varepsilon_{it}$$

The GAM approach allows the data to determine the shape of the relationship between social housing share and house prices, without imposing a linear or segmented structure in advance. This makes it particularly suitable for detecting gradual non-linear patterns, such as flat segments, steep declines, or other irregularities in the price response across the distribution of (S_{nt}). As noted by Wood (2017), GAMs are well suited for exploratory modelling when the functional form is unknown but believed to be smooth.

4.4.2 Algorithmic and Manual Piecewise Linear Model (PLM)

To test for the presence of a tipping point in the relationship between social housing share and house prices, a piecewise linear model (PLM) is estimated. The model uses the same underlying specification as the GAM, but replaces the smooth function $f\ (S_{nt})$ with two linear segments, joined at an estimated breakpoint:

$$\ln(P_{it}) = \ \alpha \ + \ \beta^1 S_{nt} + \ \beta^2 (S_{nt} - \ c) \cdot \ 1(S_{nt} > \ c) + \ \boldsymbol{X}'_{it}\beta \ + \ \delta_t + \ \varepsilon_{it}$$

Here, c represents the breakpoint in the distribution of social housing share, and $(S_{nt} > \ c)$ is an indicator function that equals 1 when the share exceeds this threshold. The slope below the breakpoint is given by β^1, while β^2 captures the change in slope above the threshold.

Two different approaches are used to estimate the breakpoint c:

In the first approach, the breakpoint is estimated endogenously using the segmented R package. A starting value is provided and the algorithm searches for the best-fitting value of c within a constrained range. In this study, the search interval is restricted to [0.07, 0.30], based on patterns observed in the LOESS and GAM plots. This constraint helps avoid implausible estimates at the tails of the distribution, where data are sparse and model fit is less reliable.

In the second approach, a transparent grid search is conducted over the same interval. A series of candidate values for ccc is defined (e.g. in 0.005 increments), and for each value, the model is re-estimated. The estimated slope change β^2 and model fit (adjusted R^2) are then recorded. This method allows for visual inspection of how model performance varies across breakpoints and avoids reliance on algorithmic optimisation criteria.

Both methods use the same functional form and regression structure. The advantage of the algorithmic method is efficiency and formal inference, while the grid search offers transparency and helps validate the stability of estimated breakpoints.

In line with the definition introduced in the literature review (Chapter 2.1), a breakpoint is interpreted as a tipping point only if the estimated change in slope (β^2) is both statistically significant and substantively large. This means that the shift above the threshold must not only be unlikely to occur by chance (typically $p < 0.05$) but must also indicate a meaningful departure from the slope below the breakpoint. In practice, this involves examining the magnitude and direction of the slope change and assessing whether it suggests an abrupt shift in the price trend, rather than a continuation of a smooth pattern. This distinction is important to avoid labelling every statistically significant kink in the relationship as a tipping point.

While less flexible than the GAM, the PLM provides a more interpretable structure, making it easier to assess whether a distinct threshold exists – and if so, how the slope of the relationship changes above and below that point. This makes the PLM a natural tool for detecting tipping point dynamics in empirical settings (Card et al., 2008).

4.4.3 Two-Way-Fixed-Effects Model (PLM)

Neighbourhood fixed effects are excluded from the baseline to retain variation in the key explanatory variable (neighbourhood-level share of social housing) both across space and over time.

However, this raises concerns about potential omitted variable bias if unobserved, time-invariant neighbourhood characteristics are correlated with both social housing share and house prices. To address this, a robustness model is estimated with neighbourhood fixed effects to absorb all time-invariant neighbourhood traits, though at the cost of reducing variation in S_{nt}. This model is specified as:

$$\ln(P_{it}) = \alpha + \beta^1 S_{nt} + \beta^2 (S_{nt} - c) \cdot 1(S_{nt} > c) + X_{it}' \beta + \delta_t + \gamma_n + \varepsilon_{it}$$

4.4.4 Two-stage Hedonic model (GAM)

A two-stage hedonic GAM is estimated to test whether the observed relationship between social housing share and house prices is driven by differences in housing structure. This is important because areas with higher or lower shares of social housing may also differ systematically in dwelling size, type, or age, which could confound the results. The model proceeds in two steps. In the first stage, log prices are regressed on structural housing characteristics and year using a standard linear model:

$$\ln(P_{it}) = \alpha + X_{it}' \beta + \delta_t + \varepsilon_{it}$$

In the second stage, these residuals are modelled as a smooth function of neighbourhood-level social housing share using a GAM. This allows for a flexible non-linear fit, while isolating the price component that is potentially attributable to local composition effects.

$$\varepsilon^{residual}{}_{it} = s(S_{nt}) + \eta_{it}$$

By comparing the resulting curve to the baseline GAM, it becomes possible to assess whether the observed pattern reflects genuine neighbourhood differences or is largely an artefact of structural variation.

4.4.5 Lagged models (GAM & PLM)

Lagged models are estimated to test whether the effect of social housing share on house prices occurs with a delay. This may be the case if housing market participants adjust their expectations gradually, or if changes in neighbourhood composition take time to be reflected in transaction prices. For example, the announcement or completion of new social housing may influence perceptions of an area only after some time, especially if it triggers broader changes in reputation, demand, or local amenities.

To account for such delayed responses, both the Generalised Additive Model (GAM) and the Piecewise Linear Model (PLM) are re-estimated using lagged values of the social housing share. Specifically, lags of 2 and 4 years are chosen. These intervals reflect plausible delays in local market adjustment, while still preserving sufficient data coverage across the time window (2011–2023). Shorter lags may be too immediate to capture adjustment dynamics, while longer lags would result in excessive data loss and less robust estimates.

The lagged GAM is specified as:

$$\ln(P_{it}) = \alpha + f\left(S_{\{n,t-k\}}\right) + X_{it}'\beta + \delta_t + \varepsilon_{it}$$

The lagged PLM is specified as:

$$\ln(P_{it}) = \alpha + \beta^1 S_{\{n,t-k\}} + \beta^2\left(S_{\{n,t-k\}} - c\right) \cdot 1\left(S_{\{n,t-k\}} > c\right) + X_{it}'\beta + \delta_t + \varepsilon_{it}$$

Both versions use the same controls and fixed effects structure as the corresponding non-lagged models, which ensures consistency in estimation.

4.4.6 Linear Mundlak model

A Mundlak-style linear model is also estimated to distinguish between within-neighbourhood and between-neighbourhood variation in the effect of social housing share on house prices. It includes neighbourhood means of the time-varying variable alongside the fixed effects. This approach addresses the possibility that part of the observed relationship is driven by long-standing differences between neighbourhoods, rather than by changes within them over time (Wooldridge, 2021).

The model includes both the time-varying neighbourhood-level share of social housing, S_{nt}, and the time-invariant neighbourhood mean of this variable, $\overline{S}_n$. The specification is:

$$\ln(P_{it}) = \alpha + \beta^1 S_{nt} + \beta^2 \overline{S}_n + X_{it}'\beta + \delta_t + \varepsilon_{it}$$

Here, β^1 captures the effect of changes in social housing share within neighbourhoods over time, while β^2 captures between-neighbourhood differences in long-run average social housing share. All models include the same structural controls and year fixed effects as in the baseline.

To provide a direct benchmark, a standard pooled OLS model is also estimated, using the same specification but without the neighbourhood means:

$$\ln(P_{it}) = \alpha + \beta S_{nt} + X'_{it}\beta + \delta_t + \varepsilon_{it}$$

This pooled model captures the total effect of S_{nt}without distinguishing whether the variation occurs within or between neighbourhoods. Comparing the two models helps assess how much of the observed relationship is driven by long-run spatial differences versus local change over time.

4.4.7 Local Polynomial Discontinuity model (RDrobust)

To test for local discontinuities in the relationship between social housing share and house prices, the RDrobust package is used to estimate local linear regressions around specific cutoffs. This method focuses on whether there is a statistically significant jump in log house prices at a given threshold of social housing share, without modelling the full distribution. The functional form estimated around cutoff ccc is:

$$\ln(P_{it}) = \alpha + \tau \cdot \mathbf{1}(S_{nt} > c) + f(S_{nt} - c) + X'_{it}\beta + \varepsilon_{it}$$

Where $\mathbf{1}(S_{nt} > c)$ is an indicator for whether the social housing share exceeds the threshold, τ captures the size of the discontinuity and $f(S_{nt} - c)$ represents a local polynomial fit on either side of the cutoff. The method is applied at two thresholds: 14.5% and 19%, based on breakpoints identified in the PLM models. While this approach does not estimate slopes or overall fit statistics, it offers local evidence on whether the relationship changes abruptly at key points in the distribution.

4.4.8 Estimation details

All models are estimated in RStudio using R version 4.4.1. The main dataset is constructed by merging housing transaction data from NVM (2000–2023) with manually compiled social housing share data (2011–2024), resulting in a cleaned regression sample of 104,716 transactions. Data were pre-processed using `dplyr`, `data.table`, `readr`, and `readxl`, with custom manual corrections to neighbourhood codes to ensure consistency across sources.

For the Generalised Additive Models (GAMs), the `mgcv` and `gratia` packages were used. Smooths were estimated using penalised thin plate regression splines, and smooth confidence intervals were visualised using `ggplot2`. Piecewise Linear Models (PLMs) were estimated using the `segmented` package, with breakpoints estimated via constrained search within a plausible range (typically 7–30%) to avoid overfitting in sparse tails of the distribution. Convergence was ensured by using fixed starting values and setting the number of bootstrap iterations to zero for stability.

Mundlak-style models and all linear regressions were estimated using base R's `lm()` function. The pooled OLS benchmark and the Mundlak model were implemented in parallel to isolate within- and between-neighbourhood effects, with coefficients and confidence intervals compared graphically.

Lagged models were estimated using both PLM and GAM specifications, with 2-year and 4-year lags of the neighbourhood-level social housing share. Lag variables were created using `dplyr::lag()` grouped by neighbourhood and sorted by year.

For the first stage of the hedonic residual model, log house prices were regressed on structural controls (living area, room count, dwelling type, construction period), and the residuals were extracted. These were used as the dependent variable in GAMs (`mgcv`) to assess the effect of social housing share net of housing characteristics.

A regression discontinuity-style sensitivity check was conducted using the `rdrobust` package, focusing on a potential discontinuity around the estimated breakpoint from the PLM model. While not based on a strict running variable design, this local fit offers an additional check for sharp changes in slope and model robustness near the threshold.

All models include year fixed effects via `factor(year)` to control for citywide housing market trends and macroeconomic shocks. Where relevant, neighbourhood fixed effects were added using `factor(neighbourhood_code)`.

Standard errors are not explicitly clustered but the model structure accounts for grouped data through fixed effects where applicable. Data were trimmed to exclude neighbourhoods with over 90% social housing, and observations with missing values on key controls were removed during pre-processing. Visualisations were generated using `ggplot2`, and tables for the thesis were produced using `xtable`, `modelsummary`, and `flextable`.

4.5 Identification and assumptions

While this study does not aim to establish a casual effect of social housing share on house prices, it is still important to consider where the identifying variation comes from and under what conditions the estimated relationship can be interpreted as meaningful. The analysis relies on pooled cross-sectional transaction data. Variation in the neighbourhood-level share of social housing occurs both across space (between neighbourhoods) and time (within neighbourhoods). This mixed structure is central to the identification strategy: cross-sectional variation helps to reveal broad patterns across the city, while temporal variation within neighbourhoods allows the models to pick up local dynamics.

A key concern in this type of analysis is that the neighbourhood share of social housing is unlikely to be strictly exogenous. It may correlate with unobserved neighbourhood characteristics that also affect house prices, such as crime levels, access to amenities, or average income levels. If these unobserved factors are stable over time, they may lead to omitted variable bias in models without fixed effects. In addition, anticipation effects may play a role: house prices might respond to expected changes in social housing policy or redevelopment, even before the actual housing stock changes. This would further complicate causal interpretation.

Unlike settings where a clear external shock or natural experiment provides quasi-random variation, this study does not rely on an instrument or treatment event. As such, endogeneity concerns remain and no causal claims are made. The goal is not to estimate the causal effect of increasing social housing share, but rather to detect robust non-linear patterns in the observed correlation between social housing and house prices. That distinction is important: the focus is on describing the shape of the relationship, not on predicting what would happen if a policy were to increase or decrease the share of social housing.

To assess whether the observed patterns are meaningful and not just driven by noise or bias, several strategies are used. One approach is to include neighbourhood fixed effects, which control for all local characteristics that do not change over time. This helps isolate how house prices shift within the same neighbourhood when the share of social housing changes. Another approach is the Mundlak model, which separates long-term differences between neighbourhoods from changes that happen locally over time. This makes it possible to observe whether the relationship is mainly shaped by structural differences across the city or by local developments in the housing stock. Lastly, lagged models are used to examine whether house prices respond with a delay, capturing possible gradual adjustment dynamics rather than immediate responses.

This study does not make causal claims, as it does not rely on exogenous variation or a clearly defined treatment. However, establishing causality is not the objective. The aim is to examine whether the correlation between the neighbourhood share of social housing and local house prices is linear or non-linear, and in case of the latter, whether a clear tipping point can be identified. The empirical strategy is well aligned with this goal. By combining flexible estimation techniques, segmented models, fixed effects, lagged structures, and within-between decompositions, the analysis is designed to detect consistent patterns in the data and to test their robustness across a range of plausible assumptions.

5. Results

This chapter presents the main results of the analysis. A smooth GAM shows the overall shape, and two PLMs detect and compare slope changes. A summary table brings the core results together, including breakpoints, slopes, and fit statistics, to make it easier to compare models. The chapter ends with a selection of robustness checks that test how stable the results are across different model choices. Interpretation is kept to a minimum at this stage and left to the discussion chapter. Strong interpretations are deliberately avoided at this stage and will follow in the discussion chapter.

5.1 Regression results baseline models

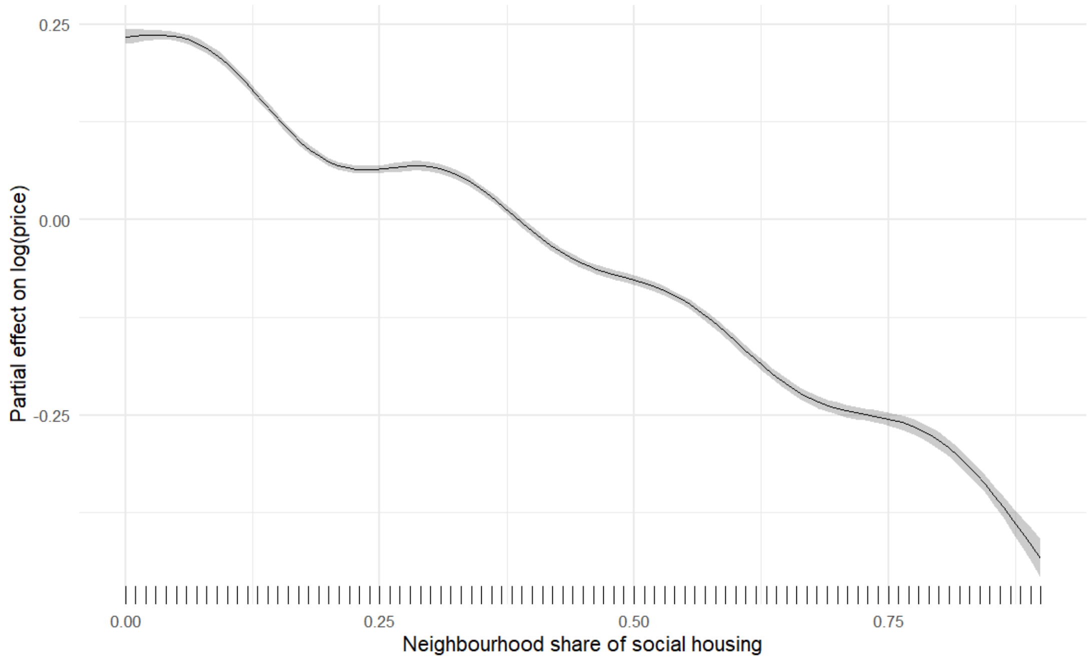
This section presents the results of the baseline models. It begins with the Generalized Additive Model (GAM) to capture the full shape of the association, followed by the two Piecewise Linear Models (PLM) that estimate changes in slope at specific breakpoints. Together, these models provide a first step to an answer to the central question of whether there is a tipping point in the relationship.

5.1.1 Baseline: Generalized additive model (GAM)

The baseline GAM model allows the effect of social housing share to vary smoothly, rather than assuming a fixed linear relationship. Figure 5.1 shows the estimated smooth function $f(S_{nt})$, based on the trimmed sample. The y-axis represents the partial effect of social housing share on the log of house prices, holding all other variables constant.

The model shows a clear and consistent downward trend: prices are relatively stable up to about 7-8%, then decline steeply between 8% and 20%, after which the slope levels off briefly. Beyond 30%, the decline continues more gradually across the upper half of the distribution. While there is no sharp kink or inflection that would suggest a tipping point, the pattern is for from linear. The smooth function captures a steady change in slope across the distribution, with the most pronounced price effects emerging in the lower to middle range of social housing shares. This supports the use of flexible, non-linear methods in the rest of the analysis.

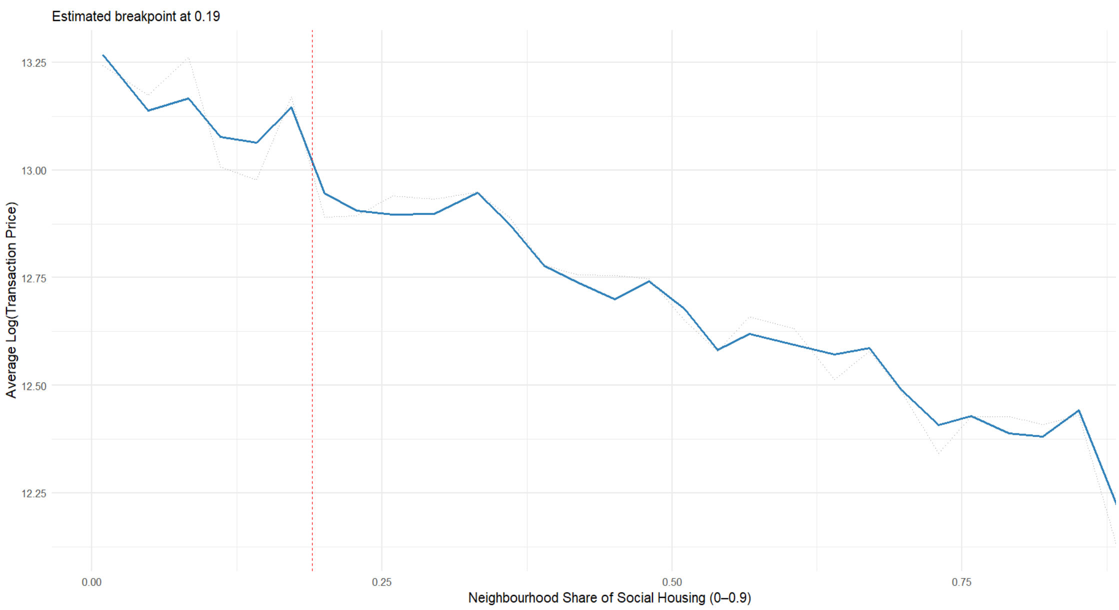
Figure 5.1: Baseline GAM: Smooth effect of social housing share



5.1.2 Baseline: Piecewise linear models (PLM)

Two piecewise linear models (PLMs) are estimated: one using an algorithmic breakpoint search and one based on a manual grid search. Both use a constrained breakpoint range of 0.07 to 0.3, motivated by the GAM curve, which suggested relative stability below 7–8% social housing and a levelling off around 20–30%. This range helps avoid implausible breakpoints at the tails.

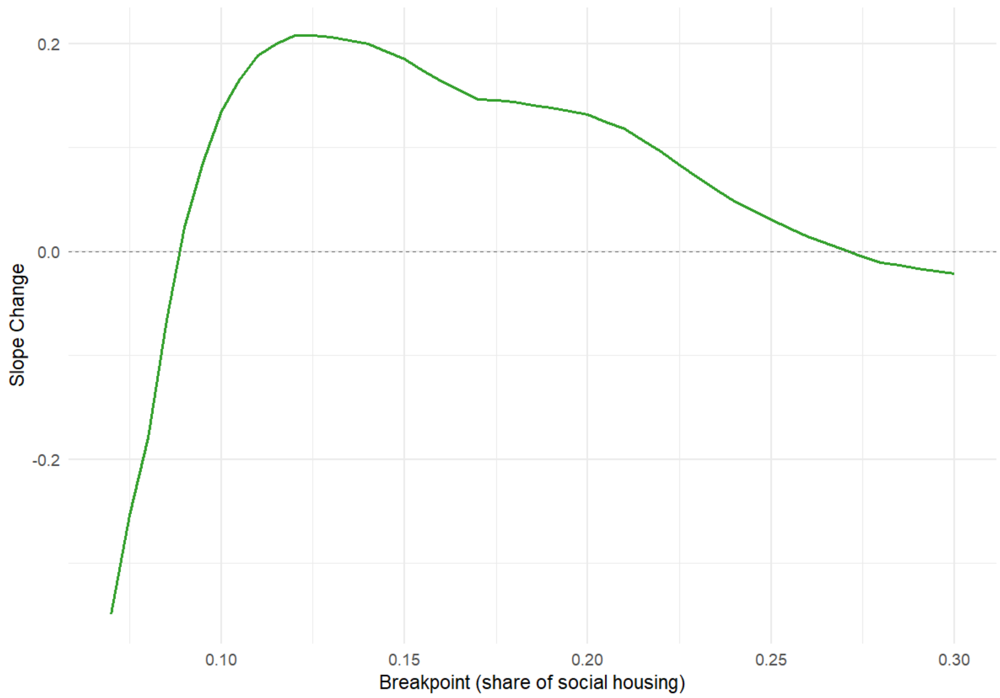
Figure 5.2: Baseline Plotted Piecewise Linear Model (PLM) with Algorithmic Breakpoint [Search Range: 0.07–0.30]



The algorithmic PLM, shown in Figure 5.2, identifies a statistically significant breakpoint at 19%. Below this point, the slope is sharply negative (–0.754), while above it the decline continues but becomes slightly less steep. The slope change ($\Delta\beta = 0.097$) is statistically significant, suggesting a modest flattening rather than a reversal. The model fits the data well (Adj. $R^2 = 0.642$; RMSE = 0.350), and supports the idea of a segmented, but still broadly downward, relationship.

To complement this, a manual grid search was run over the same range. This approach estimates the model repeatedly for different breakpoints and tracks both model fit and slope changes. As shown in Figure 5.3, the best-fitting specification is found at 14.5%, with a slope below the threshold of –0.833 and a slope change of 0.192. This puts the breakpoint somewhat earlier than the algorithmic model and implies a more gradual levelling off that starts already at moderate social housing shares.

Figure 5.3: Baseline Piecewise Linear Model (PLM) with Manual Breakpoint Grid Search [Search Range: 0.07–0.30]



Together, the two PLMs confirm the non-linear pattern found in the GAM, with a clear change in slope at moderate social housing shares. The fact that the estimated breakpoints differ – 19% in the algorithmic model and 14.5% in the manual grid search – suggests some uncertainty about the exact location of this change. This underlines the need for additional models to assess the stability and robustness of the observed structure.

5.2 Robustness and comparative results

5.2.1 Overview of estimated relationships

Table 5.1 brings together the key results from all baseline and selected robustness models. It shows the estimated breakpoint (if identified), the slopes below and above that point and model fit statistics like adjusted R^2 and RMSE. For models that do not include a breakpoint, such as the GAM and Mundlak specification, the table reports the relevant coefficients or discontinuity estimates.

Table 5.1: Summary of Baseline and Robustness Model Estimates

Model	Estimation technique	Breakpoint	Slope below (β_1)	Slope change ($\Delta\beta$)	Slope above (β_2)	Adj. R^2	RMSE	β within	β between	Discont inuity
Baseline GAM	GAM	—	—	—	—	0.644	—	—	—	—
Baseline Algorithmic	PLM	0.190	- 0.754***	0.097**	- 0.657***	0.642	0.350	—	—	—
Baseline MGS	PLM	0.145	- 0.833***	0.192***	- 0.641***	0.643	0.350	—	—	—
TWFE	PLM (FE)	0.183	- 0.183	- 0.758***	- 0.575***	0.808	0.256	—	—	—
2Stage Hedonic	GAM (residuals)	—	—	—	—	—	—	—	—	—
2-year Lagged	PLM	0.190	- 0.717***	0.098**	- 0.619***	0.642	0.349	—	—	—
4-year Lagged	PLM	0.190	- 0.705***	0.094*	- 0.611***	0.642	0.349	—	—	—
2-year Lagged	GAM	—	—	—	—	0.643	—	—	—	—
4-year Lagged	GAM	—	—	—	—	0.643	—	—	—	—
Mundlak	Linear	—	—	—	—	0.633	—	- 0.609***	- 0.051	—
RDrobust @14.5%	Local Polynomial	0.145 (set)	—	—	—	—	—	—	—	+ 0.369***
RDrobust @19%	Local Polynomial	0.190 (set)	—	—	—	—	—	—	—	- 0.262***

Notes: Significance levels: $p < 0.1$ (*), $p < 0.05$ (**), $p < 0.01$ (***). Significance stats are shown only for coefficient estimates: slope below (β_1), slope change ($\Delta\beta$), slope above (β_2), β within, β between and discontinuities. Goodness-of-fit metrics (e.g. Adjusted R^2, RMSE) are not subject to significance testing.

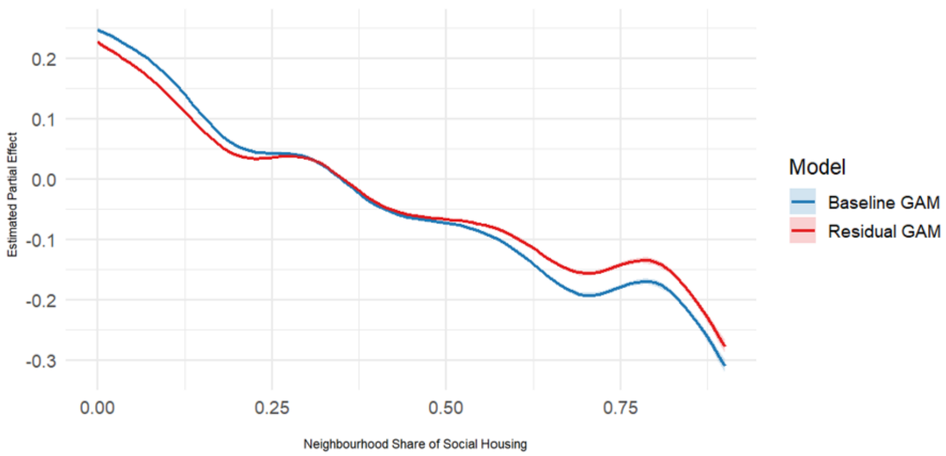
The results confirm a broadly consistent pattern: house prices decline as the neighbourhood share of social housing increases, but not in a purely linear way. The GAM reveals a smooth, non-linear curve with the steepest decline between roughly 10% and 30%. Both piecewise linear models detect a structural shift in this range. The algorithmic PLM places the breakpoint at 19%, with a strong negative slope below (–0.754***) and a slightly less steep decline above (–0.657***). The manual grid search, in contrast, identifies a breakpoint at 14.5%, with similarly significant slope estimates (–0.833*** below and –0.641*** above), suggesting the curve may start flattening a bit earlier. Despite these differences, both models indicate a significant change in slope, supporting the interpretation of a segmented but continuous relationship.

5.2.2 Results of robustness models

5.2.2.1 Two-Stage Hedonic GAM

The robustness checks in the table largely confirm the segmented but continuous pattern. The 2-stage hedonic GAM helps isolate the effect of neighbourhood social housing share from differences in housing structure. In this approach, housing characteristics and year are first used to predict log prices in a linear regression; the residuals are then modelled as a smooth function of social housing share. The resulting curve, shown in Figure 5.4, appears similar to the baseline GAM; both show a gradual, non-linear decline in price levels as social housing share increases. This similarity suggests that the observed relationship is not simply driven by differences in dwelling size, type or age. While the residual line is somewhat flatter than the baseline at the extremes, the two curves are nearly identical between 15% and 60%. This reinforces the idea that the main price pattern reflects a genuine neighbourhood-level association, rather than a compositional effect.

Figure 5.4: Estimated partial effect of social housing share on log(price) in the baseline and residual GAM



5.2.2.2 Neighbourhood Fixed Effects PLM

The fixed effects PLM provides a complementary robustness check by controlling for unobserved, time-invariant neighbourhood characteristics. Similar to the baseline PLM, it identifies a breakpoint around 18.3%, confirming the same structural location. However, the slope pattern shifts: the slope below the breakpoint becomes weakly positive and statistically insignificant (–0.183), while the decline above the threshold remains strong and highly significant (–0.575***). This suggests that, once differences between neighbourhoods are held constant, the main price effects emerge at higher levels of social housing share. The fixed effects model also achieves a better overall fit (adjusted $R^2 = 0.808$), though the fitted relationship becomes more sensitive to local variation due to the limited within-neighbourhood movement over time. Together, these two robustness checks show that the non-linear pattern holds up when accounting for housing structure and neighbourhood differences, with the strongest price effects appearing above the breakpoint.

5.2.2.3 Lagged Models

The lagged PLMs (2 and 4 years) produce almost identical slopes and breakpoints to the baseline model. All slope estimates remain highly significant, including the post-breakpoint effects (–0.619*** and –0.611***), reinforcing the idea that the segmented pattern holds over time. The breakpoint itself also stays fixed at 19%, suggesting no delayed structural change in the relationship.

The lagged GAMs support this conclusion. Their smooth functions are visually indistinguishable from the baseline, with the same steep decline between roughly 10% and 25% and a slight flattening beyond. The near-perfect overlap in the plots implies that the observed non-linear relationship is not sensitive to short-term timing effects. This stability is likely due to the slow-moving nature of social housing shares across years, which means that lagging the variable does little to alter the estimated effect.

5.2.2.4 Mundlak Linear Model

The Mundlak model offers a complementary perspective by decomposing the estimated effect into within- and between-neighbourhood components. It reveals a strong and statistically significant within-neighbourhood effect (–0.609***), which suggests that as the social housing share increases within the same neighbourhood over time, house prices tend to decline. In contrast, the between-neighbourhood effect is small (–0.051) and statistically insignificant, which indicates that structural differences across neighbourhoods explain little of the variation once other factors are controlled for.

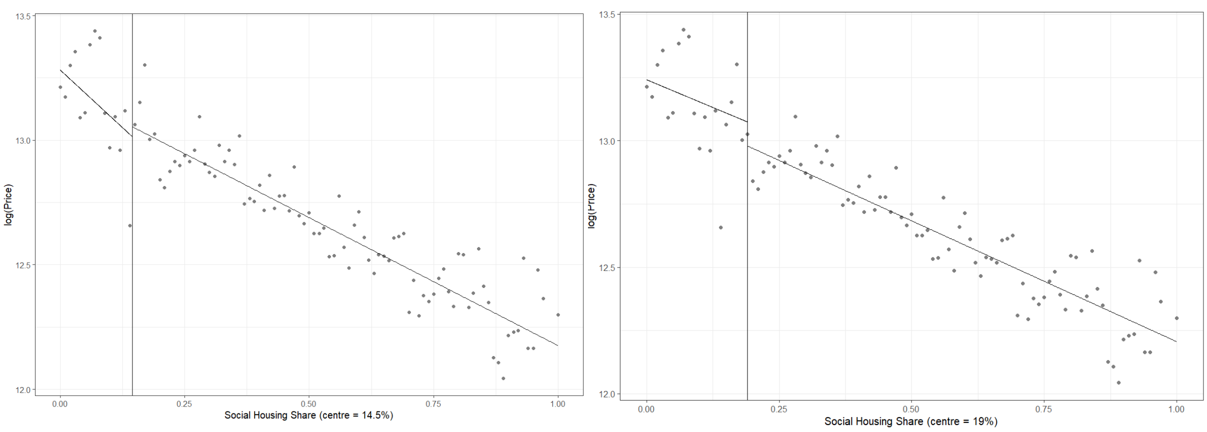
At first glance, this might seem at odds with the lagged models, which showed very stable results when the social housing share was shifted by two or four years. However, the findings are not necessarily contradictory. The Mundlak model shows that within-area variation matters, but it does not imply that these changes are large or sudden. In fact, because changes in social housing share tend to be gradual, even a 2–4 year lag does not meaningfully alter the estimated relationship. What the Mundlak model adds is that these gradual, small, local shifts are still the primary driver of the observed price patterns, rather than persistent differences between neighbourhoods.

This strengthens the interpretation that the relationship between social housing and house prices is not just robust over time, but also largely shaped by local dynamics.

5.2.2.5 Local Discontinuity tests (RDrobust)

Finally, the discontinuity tests – the RDrobust models – provide a local check on the relationship around the two key breakpoints identified earlier. Figure 5.5 shows that, at the 19% cutoff (right side), a clear downward shift is visible in the RD plot, and the estimated discontinuity (–0.262***) aligns well with the slope change found in the algorithmic PLM. This supports the idea that the price decline becomes more pronounced beyond this point. At 14.5%, by contrast, the RD plot shows a small but statistically significant upward jump (+0.369***), which does not match the flattening pattern suggested by the manual PLM. This inconsistency indicates that the price effect at this lower threshold may be more sensitive to local variation or noise.

Figure 5.5: RD plot at 14.5% (left) and 19% (right) social housing share, showing a small but statistically significant discontinuity in log house prices at the threshold



5.3 Summary of results

The analysis consistently shows that neighbourhood social housing share is negatively associated with house prices, but the relationship is far from linear. Prices remain relatively stable at low levels of social housing, then begin to decline more sharply between roughly 10% and 30%, before flattening out again toward the upper end of the distribution. This non-linear pattern emerges clearly in the GAM and is reinforced by the piecewise linear models, which detect a significant change in slope around 14.5% (manual grid search) and 19% (algorithmic). While these breakpoints differ, both indicate that the price effect intensifies over this middle range, without suggesting a sharp tipping point.

The robustness checks largely confirm and refine this picture. Models with neighbourhood fixed effects show that the price decline is concentrated above the breakpoint, once time-invariant neighbourhood characteristics are accounted for. The 2-stage hedonic GAM reveals that the pattern is not simply driven by housing structure, while the Mundlak model highlights that it is mostly local, within-neighbourhood changes in social housing share that matter for price variation. The lagged models support this, showing near-identical results with 2- and 4-year lags, which suggests that effects are stable and not driven by short-term timing. Finally, the local discontinuity tests support the interpretation that 19% marks a meaningful, albeit modest, structural shift in the relationship. The RDrobust model at this cutoff shows a clear and statistically significant downward jump in log prices, consistent with the breakpoint identified in both the algorithmic PLM and the fixed effects specification. In contrast, the test at 14.5% reveals a small upward discontinuity, which does not align with the flattening observed in the manual PLM. This suggests that 14.5% is not a structural breakpoint in the strict sense, but rather reflects a more gradual levelling off.

Overall, the robustness checks reinforce that the estimated breakpoint captures a genuine shift in slope. While not a sharp tipping point, it marks a consistent inflection in the downward trend. The association between neighbourhood social housing share and local house prices in Amsterdam neighbourhoods can best be described as a gradual decline with a subtle, but notable change around 19% social housing share.

6. Discussion

This chapter discusses and interprets the empirical findings presented in the previous chapter. Section 6.1 focuses on the main results and their meaning. Section 6.2 discusses the limitations of the study. Section 6.3 reflects on the broader research implications and potential directions for future work. Finally, Section 6.4 outlines the policy relevance of the findings and draws connections to ongoing debates in urban housing policy.

6.1 Findings and interpretations

Across all models, the relationship between social housing share and house prices is non-linear, segmented and consistently downward-sloping. Prices remain relatively stable at low shares, decline more steeply in the mid-range, and then flatten at higher shares. This pattern is consistent across flexible and segmented regression models and mirrors gradual transitions in neighbourhood composition and housing dynamics found in earlier work (Bruch and Mare 2006; Rosenthal and Ross, 2015).

Visual exploration using LOESS and binned scatterplots reveals a broadly negative relationship, but not a linear one. The relationship varies more at low social housing shares, declines more steadily across the middle of the distribution (roughly 10 to 70 percent), and appears to level off toward the upper end. At very low levels of social housing, the relationship with house prices is negative but appears to be irregular. This may indicate that small increases in social housing share do not consistently influence prices when the overall presence is limited. As the share increases, the pattern becomes more stable and more clearly downward sloping. This pattern is consistent with gradual sorting effects and local perception mechanisms (Holvandus and Leetmaa, 2016).

The baseline regression models confirm the patterns observed in the visual analysis and provide stronger evidence of a non-linear relationship between neighbourhood social housing share and house prices. In this thesis, a tipping point is defined as a discrete and statistically significant change in slope that is also substantively large, indicating a sharp break in the price trend (Card et al., 2008; Malone, 2020). While the PLM identifies a statistically significant breakpoint around 19 percent, the associated change in slope is small and does not reflect an economically meaningful shift. This interpretation is supported by the GAM, which shows a smooth curve without any abrupt changes in direction. These findings align with Bruch and Mare's (2006) argument that tipping behaviour requires strong behavioural thresholds, which may not exist in this setting. Instead, the evidence points to a gradual, continuous variation in the effect across the distribution.

The manual grid search placed the breakpoint slightly earlier, at 14.5%, but this does not contradict the main result. The difference is expected in a setting where no sharp or dominant breakpoint exists. Because the underlying trend changes gradually, not abruptly, different estimation approaches are likely to detect slightly different points of inflection. Rather than revealing inconsistency, this variation confirms the absence of a strong tipping dynamic and reinforces the interpretation of a segmented but continuous relationship. Similar findings are reported by Tomal and Rahman (2021), who show that estimated breakpoints in housing markets can vary notably across methods, especially when no clear structural shift is present.

Tipping points are often interpreted as signals of self-reinforcing feedback effects, where a shift in social composition leads to further declines in demand or neighbourhood reputation (Saiz and Wachter, 2011; Ong, 2017). If such feedback mechanisms were present in the relationship between social housing and house prices in Amsterdam, one would expect a sharp acceleration in the price decline beyond a certain threshold. Yet the results do not support this: although a clear change in slope is found around 19% in most models, this change is modest and the decline continues gradually rather than intensifying.

The robustness of the findings is further supported by a range of alternative models. The neighbourhood fixed effects model confirms a breakpoint in the same range and reproduces the strong negative slope above it. While the slope below the threshold becomes weaker and statistically insignificant once neighbourhood fixed effects are included, the overall segmented structure remains intact. This suggests that the downward price effect is robust, especially in the upper part of the distribution, even if the full slope pattern varies slightly across specifications.

The lagged models show that the relationship between social housing share and house prices remains stable even when the variable is shifted back by two or four years. This suggests that the pattern is not driven by short-term market adjustments but reflects more structural dynamics. Because social housing shares tend to change slowly over time, lagged values remain close to contemporaneous ones, which explains why the estimated effect hardly shifts. This contrasts with studies such as Autor et al. (2014), who interpret lagged effects as signs of delayed responses in housing markets. In this context, the stability of the lagged models points instead to gradual changes in neighbourhood composition as the main driver of the observed relationship.

This interpretation aligns with earlier research on neighbourhood transformation, which shows how price changes can accumulate gradually without being triggered by large shocks. For instance, studies by Guerrieri et al. (2013) and Ellen and O'Regan (2011) describe how house price appreciation unfolds unevenly within cities, often through subtle shifts in demand, reputation or composition. In a similar way, the results here suggest that social housing share acts as a slow-moving driver of neighbourhood change, influencing prices in a continuous but locally grounded way. Rather than reflecting threshold effects or sharp transitions, the relationship appears to capture how small, incremental changes in tenure mix can shape housing demand over time.

So rather than reflecting static neighbourhood differences, the observed pattern appears to capture dynamic local shifts as social housing share increases or decreases. Since this study does not measure causal effects, the within-neighbourhood changes in social housing share may still coincide with other factors that vary over time, such as local investment, demographic change or housing improvements, which can also influence prices. The observed correlation may therefore reflect a broader set of neighbourhood developments rather than the effect of social housing in isolation. This is in line with other studies that aim to detect structural patterns without claiming causal identification (Albright et al., 2013).

However, the lack of causality does not mean it's not possible to interpret the mechanisms behind this correlation. The gradual relationship that is identified in this study may reflect how social housing becomes more visible as its share increases. At low levels, social housing may be less noticeable or integrated into mixed-tenure environments. As the share grows, it may become more closely associated with reputational factors that shape buyer perceptions, in line with the mechanism described by Diamond and McQuade (2019) and Permentier et al. (2009). The latter study shows that perceived neighbourhood reputation in the Netherlands is strongly shaped by the presence of social housing and its symbolic meanings, including notions of disorder, low status and social problems, even when objective conditions are similar. These reputational effects influence not only residents' satisfaction but also the willingness of outsiders to move in, reinforcing demand-side responses to visible social housing concentrations. This could explain why the strongest negative slopes are found in the middle of the distribution, where social housing is present in meaningful quantities but not yet fully dominant.

Neighbourhood change in Amsterdam and other European cities tends to unfold more gradually than in many American contexts, where tipping point dynamics are more commonly studied. In the U.S., sharp thresholds in racial or income composition have been linked to sudden shifts in housing demand (Card et al. 2008), but European urban change is more often shaped by institutional factors, long-term planning and slower demographic turnover (Ong, 2017; Zorlu and Latten, 2009). In the Netherlands, neighbourhood transformation is driven by tenure restructuring, targeted investment, and state-regulated housing policy. Van Gent and Hochstenbach (2019) highlight how the reduction of social housing and the rise of market-oriented tenure forms proceed incrementally, altering neighbourhood composition without triggering abrupt behavioural responses. This context helps explain why the relationship observed in this study is smooth and continuous, rather than marked by tipping dynamics.

Finally, the findings also shed light on how the social housing share shapes neighbourhood attractiveness. The gradual (non-tipping) decline in prices suggests that desirability is not suddenly 'lost' at specific thresholds, but instead erodes incrementally as social housing becomes more prominent. At low shares (below 15 percent), social housing may blend into mixed-tenure environments with little effect on demand. In the mid-range (15 to 65 percent), reputational effects likely intensify leading to steadier declines in desirability. This is consistent with Permentier et al.'s (2009) findings on neighbourhood stigma in the Dutch context. At very high shares (above 70 percent), prices tend to level off, possibly because buyers who prioritise affordability over status self-select into these areas. This suggests that how social housing affects desirability depends on the local context and that good design, integration and spatial distribution may be more important than sticking to fixed quotas.

Overall, the empirical pattern reflects a non-linear, segmented but continuous association, where social housing share increasingly aligns with neighbourhood characteristics that influence housing demand. This interpretation is consistent with both the visual analysis and the regression results and supports the idea that neighbourhood change in Amsterdam is shaped by gradual, rather than by threshold-based, dynamics.

6.2 Limitations

While the analysis in this study is carefully designed and supported by extensive data and robustness checks, several limitations should be acknowledged.

First, there are limitations related to data availability and variable construction, as discussed in Subchapter 3.3. The social housing dataset was constructed manually and only covers the period from 2011 onward, meaning earlier transactions are excluded from the main analysis. Neighbourhood boundaries are held constant using 2022 CBS codes, which ensures consistency but may overlook minor reclassifications over time – though the number of neighbourhoods remained largely stable. Several housing characteristics, such as plot size and gross volume, were excluded due to high rates of missing values. While their omission may leave out some variation in dwelling quality, the remaining variables are sufficient to control for key structural differences. Lastly, the process of linking neighbourhood names to CBS codes involved some manual corrections, which may have introduced minor imprecision despite careful documentation.

A second limitation is the inability to identify formerly social housing units that have been sold to private buyers. These are treated as regular transactions, which may slightly distort the relationship between social housing share and price if such homes differ systematically from other private dwellings. However, this is unlikely to affect the overall patterns in a substantial way (see e.g. Buurma-Olsen et al., 2025).

Third, although the Mundlak model shows that within-neighbourhood changes in social housing share are strongly associated with house prices, the exact underlying mechanisms remain unobservable in this dataset. Changes in the social housing share could reflect new construction, demolition, tenure conversion, or administrative reclassification, but these processes are not directly observed. As a result, the observed variation may reflect a mix of underlying processes, each of which could affect house prices in different ways (Blanco and Neri, 2023).

Another limitation concerns the Local Discontinuity Model. RDrobust estimates around the 14.5% threshold show an unexpected upward jump in prices, which is not supported by the other models. This inconsistency may reflect local noise, data sparsity, or the RD design's sensitivity to bandwidth and cutoff choice. It underscores the challenge of drawing strong conclusions from local polynomial fits without a sharp treatment assignment.

The spatial scale at which social housing share is measured also deserves attention. In this study, it is calculated at the neighbourhood level using CBS definitions. However, homebuyers may respond to more immediate surroundings, such as the street or block, creating a potential mismatch between the variable and the scale at which perceptions are formed. This relates to the Modifiable Areal Unit Problem (MAUP), which refers to how results can vary depending on the spatial units used (Openshaw, 1984). While some fine-grained variation may be masked, neighbourhoods remain a practical and policy-relevant level of analysis. They align with how housing data is reported and how policies are implemented. Also, the use of more than 500 neighbourhoods within Amsterdam allows for relatively small and detailed spatial units. Given the robustness of the results across multiple specifications, the risk of MAUP substantially influencing the main conclusions is limited.

The fact that this study focuses exclusively on Amsterdam limits the generalisability of its findings. Housing dynamics, policy frameworks and social housing systems vary widely across cities and countries (Whitehead and Scanlon, 2007), meaning the results may not translate directly to other urban contexts. However, this narrow geographic scope also offers distinct advantages. Amsterdam provides a high degree of between-neighbourhood variation in social housing share, along with detailed transaction data over a long period. This allows for a more precise and context-rich analysis than would be possible in a broader cross-city study. While the findings may not apply universally, they offer robust insights into the structure of Amsterdam's housing market and can inform similar research in other urban settings.

These limitations set important boundaries on how the results should be interpreted, particularly in terms of data coverage, measurement, generalizability and unobserved factors. Even so, they do not undermine the main findings. The study remains well designed to uncover consistent and meaningful patterns in the relationship between social housing share and house prices in Amsterdam.

6.3 Research implications

The findings of this study suggest that in Amsterdam, local house prices are negatively associated with increases in the neighbourhood share of social housing, but without strong evidence of tipping point dynamics. This smoother, non-linear relationship contrasts with U.S.-based tipping point studies, where neighbourhood change often involves sharper behavioural responses to shifts in racial or income composition (Galster et al., 2001; Lyons and Loveridge, 1993). In those contexts, the spatial concentration of subsidised or low-income housing appears to trigger a threshold response in the housing market, where property values decline more sharply once a certain density of social housing is reached. This pattern, observed in U.S. cities like Denver and Minneapolis-St. Paul, contrasts with the more gradual price dynamics found in Amsterdam.

This contrast raises important questions about the institutional and cultural context of housing markets. European, in particular Dutch cities, tend to exhibit slower demographic turnover, more stable planning frameworks and a stronger role for public institutions in shaping urban development (Zorlu and Latten, 2009; Ong, 2017; Van Gent and Hochstenbach, 2019). Amsterdam's housing dynamics unfold within a relatively stable institutional and planning context, making tipping-like responses less likely. This institutional stability suggests that housing market responses to social housing concentration in European cities may follow different dynamics than those observed in more market oriented contexts. Future research could explore how national planning frameworks, tenure regulation, and the role of housing associations shape the presence or absence of tipping points and whether such behaviour is context-dependent or a more general urban phenomenon by comparing cities with different housing regimes.

Of course, institutional and cultural differences are not the only possible explanation for the difference in outcomes of tipping behaviour between Dutch and American contexts. Income inequality, regional differences and poverty rates are substantially larger in big parts of the United States than in the Netherlands (OECD, 2021). These structural differences could also help explain how social housing concentration shapes housing demand and should be explored further.

The results of the Mundlak model show that variation in social housing share within neighbourhoods over time is more strongly associated with local house prices than long term differences between neighbourhoods. This suggests that house price responses are shaped more by local developments than by static spatial inequalities, but the underlying mechanisms remain unclear. Future research could explore what drives these local changes – whether they reflect shifts in the housing stock, changes in perceived reputation, neighbourhood investment or population turnover. Combining panel data with qualitative insights or more detailed spatial measures may help to uncover why some neighbourhoods respond more strongly to changes in social housing share, and how these responses relate to broader patterns of urban transformation.

To better understand these mechanisms, future research should also address the question of causality. The patterns observed in this study reflect strong associations, but it remains unclear whether changes in social housing share actively shape house prices or whether both are influenced by other local developments. Identifying causal effects would require research designs that leverage external shocks, policy variation or institutional reforms. Such strategies could help determine whether price shifts follow directly from housing mix changes or emerge as part of broader neighbourhood transitions.

Methodologically, this study contributes to the literature by using a large, rich panel of housing transactions combined with manually compiled data on neighbourhood social housing share. This allows for a detailed and time-sensitive analysis of local housing market patterns across more than a decade. Future research could build on this by linking transaction data to survey measures, reputation indicators or qualitative insights to better understand how perceptions of social housing influence market behaviour.

These implications point to the need for more context sensitive, mechanism oriented and causally informed research on how social housing shapes local housing markets and highlight the value of combining detailed spatial data with flexible and transparent modelling strategies.

6.4 Policy implications

Beyond academic relevance, the findings of this study point to several policy implications for Amsterdam and similar cities. The identified patterns offer guidance on how social housing can be planned and integrated at the neighbourhood level.

The results of this study suggest that, from the perspective of neighbourhood desirability, there is no clear reason to maintain the share of social housing in a neighbourhood below or above a specific threshold. Although house prices are negatively associated with higher social housing shares, the relationship is gradual rather than driven by a sharp behavioural response. There is no evidence of a tipping point where market demand suddenly drops. Instead, perceptions and prices appear to shift slowly across the distribution. This points to the need for a more flexible approach to housing mix, focused less on fixed limits and more on the quality, distribution and integration of social housing.

At the same time, the negative association should not be seen as an argument against expanding or preserving social housing. House prices reflect buyer preferences and can serve as a proxy for neighbourhood desirability, but they do not include broader goals like affordability, inclusion or long term stability. A lower price level may simply reflect greater accessibility or a more diverse housing stock. From a policy perspective, the aim is not to maximise house prices, but to create neighbourhoods that are liveable and balanced. The gradual relationship suggests that social housing can be added without causing sudden shifts in market demand.

The nonlinearity of the relationship also has implications for policy. It suggests that the effect of adding social housing depends on the existing situation in a neighbourhood. In areas with very low or very high shares, the impact on house prices appears limited. In the middle range, where the decline is steeper, changes in housing mix may be more noticeable. This points to the need for a more context specific approach, where local conditions guide how and where social housing is added.

In sum, this study highlights that the relationship between social housing and neighbourhood desirability is neither fixed nor sudden. Instead of relying on rigid threshold or one size fits all strategies, effective policy should recognise the gradual nature of market responses and focus on thoughtful integration, long-term balance and locally informed decisions.

7. Conclusion and contributions

This chapter aims to answer the central question of this study and to highlight its most important empirical and methodological contributions. The main question of this thesis was:

Is the relationship between neighbourhood social housing share and local house prices in Amsterdam linear, gradually non-linear, or marked by a tipping point?

This thesis examined the shape of the relationship between social housing shares and neighbouring house prices in Amsterdam. Drawing on over 100,000 housing transactions from 2011 – 2023, combined with detailed data on social housing shares, the study explored whether this relationship is linear, gradual or market by a tipping point. The analysis shows that house prices tend to decline as the social housing share increases, but this decline unfolds gradually rather than abruptly. No sharp threshold was found beyond which prices drop disproportionately. Instead, the relationship appears to be non-linear and smooth, which suggests that neighbourhood desirability – proxied by house prices – changes incrementally with social housing composition and not through sudden shifts.

This gradual pattern was robust across all models and supported by both flexible and piecewise regression methods. A breakpoint was detected at around 19 percent, but the associated slope change was small and not substantively meaningful. The results suggest that neighbourhood change in Amsterdam is shaped by slow-moving dynamics, which is in line with existing work on gradual urban transformation. Although no causal effects were estimated, the findings offer new insight into how the composition of the housing stock relates to housing market outcomes. The smooth and continuous decline in house prices suggests that social housing shares interact with local perceptions and reputational dynamics in a gradual way.

Another important finding came from the Mundlak-model, which showed that the association was driven mainly by within-neighbourhood variation over time, rather than by long-term differences between neighbourhoods. This reinforces the interpretation that changes in social housing share reflect, and possibly contribute to, local processes of neighbourhood development (Van Ham et al., 2012).

The main contribution of this thesis lies in tracing the shape of the relationship between social housing share and local house prices in Amsterdam, using a large and detailed dataset without relying on strong functional assumptions. Rather than starting from a fixed model structure, the study builds up the analysis step by step. This makes it possible to capture gradual variation, detect slope changes and assess robustness across specifications, while staying close to the data. The empirical approach shows how non-linear relationships in spatial housing data can be investigated in a transparent and structured way.

In relation to the tipping point literature, this thesis adds nuance by focusing on housing tenure composition rather than demographic change. Whereas much of the established work identifies tipping points in response to shifts in ethnic or income composition, this study examined whether similar dynamics apply to social housing share. The absence of a sharp threshold suggests that tenure-based tipping points may not manifest in the same way as demographic ones, particularly in settings with more institutional stability and slower patterns of neighbourhood change. This contrasts with findings from several U.S.-based studies, where tipping points have been observed in response to the concentration of low-income or subsidised housing, often triggering sudden drops in local property values (Galster et al., 2001). In context of Amsterdam, however, the relationship appears smoother and more continuous. This reinforces the idea that tipping behaviour is not universal but shaped by local context, planning regimes and urban structure. Instead, the findings align with recent research that emphasises gradual transitions and context-dependent responses (Van Gent & Hochstenbach, 2019).

To conclude, while this study found that the relationship between social housing and neighbouring house prices is negative, non-linear and segmented, no economically significant tipping point was identified. This matters because it suggests that neighbourhood dynamics are not always governed by fixed thresholds or abrupt behavioural shifts, but can be influenced by gradual changes that unfold over time. For researchers, this underlines the importance of local and institutional context, and within-neighbourhood variation. For policy, it highlights that decisions about housing mix should not rely on one-size-fits-all rules or arbitrary quotas, but instead consider how local environments respond to gradual change.

8. References

Albright, Len, Elizabeth S. Derickson, and Douglas S. Massey. 2013. 'Do Affordable Housing Projects Harm Suburban Communities? Crime, Property Values, and Taxes in Mount Laurel, NJ'. *City & Community* 12 (2): 89–112. https://doi.org/10.1111/cico.12015.

Andersson, Roger, and Åsa Bråmå. 2004. 'Selective Migration in Swedish Distressed Neighbourhoods: Can Area-based Urban Policies Counteract Segregation Processes?' *Housing Studies* 19 (4): 517–39. https://doi.org/10.1080/0267303042000221945.

Autor, David H., Christopher J. Palmer, and Parag A. Pathak. 2014. "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts." *Journal of Political Economy* 122 (3): 661–717. https://doi.org/10.1086/675536.SCIRP+1IDEAS/RePEc+1

Baum-Snow, Nathaniel, and Justin Marion. 2009. "The Effects of Low Income Housing Developments on Neighborhoods." *Journal of Public Economics* 93 (5–6): 654–666. https://doi.org/10.1016/j.jpubeco.2009.01.001.

Bell, Alex, Sophie Calder-Wang, and Shusheng Zhong. 2023. *Pricing Neighborhood Amenities: A Proxy-Based Approach*. SSRN [Research Paper]. https://doi.org/10.2139/ssrn.4565093.

Bisin, Alberto, Andrea Moro, and Giorgio Topa. 2011. *The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions*. NBER Working Paper No. 17196. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w17196.

Blanco, Hector, and Lorenzo Neri. 2025. *Knocking It Down and Mixing It Up: The Impact of Public Housing Regenerations*. SSRN [Working Paper]. https://doi.org/10.2139/ssrn.5159297.

Boelhouwer, Peter. 2018. *De woningmarkteffecten van de woningwet*. Delft: Delft University of Technology, Faculty of Architecture, OTB Research Institute.

Bruch, Elizabeth E., and Robert D. Mare. 2006. 'Neighborhood Choice and Neighborhood Change'. *American Journal of Sociology* 112 (3): 667–709. https://doi.org/10.1086/507856.

Buurma-Olsen, Jennifer, Hans Koster, Jos N. Van Ommeren, and Jort Sinninghe Damsté. 2023. 'Quantifying Misallocation of Public Housing'. SSRN. https://doi.org/10.2139/ssrn.4657660.

Card, David, Alexandre Mas, and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *Quarterly Journal of Economics* 123 (1): 177–218. https://doi.org/10.1162/qjec.2008.123.1.177.

Clark, William A.V., and Maarten van Ham. 2009. "Neighbourhood Choice and Neighbourhood Reproduction." *Environment and Planning A* 41 (12): 2889–2902. https://doi.org/10.1068/a41248.

Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368): 829–36. https://doi.org/10.1080/01621459.1979.10481038.

De Vries, Sjerp, Robert A. Verheij, Peter P. Groenewegen, and Peter Spreeuwenberg. 2003. "Natural Environments—Healthy Environments? An Exploratory Analysis of the Relationship between Greenspace and Health." *Environment and Planning A* 35 (10): 1717–31. https://doi.org/10.1068/a35111.

Diamond, Rebecca, and Tim McQuade. 2019. 'Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development'. *Journal of Political Economy* 127 (3): 1063–1117.

Du Preez, Mario, and Michael Sale. 2013. "The Impact of Social Housing Developments on Nearby Property Prices: A Nelson Mandela Bay Case Study." *South African Journal of Economics* 81 (3): 451–466. https://doi.org/10.1111/j.1813-6982.2012.01337.x.

Dutch National Government. 2024. “Woontop: Concrete Afspraken voor 100.000 Nieuwe Woningen per Jaar.” December 11, 2024. https://www.rijksoverheid.nl

Favilukis, Jack, Pierre Mabille, and Stijn Van Nieuwerburgh. 2023. “Affordable Housing and City Welfare.” *Review of Economic Studies* 90 (1): 293–330. https://doi.org/10.1093/restud/rdac024.

Galster, George C., Anna M. Santiago, Robin E. Smith, and Peter A. Tatian. 2001. *Assessing Property Value Impacts of Dispersed Housing Subsidy Programs*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Galster, George C., and Sean Killen. 1995. “The Geography of Metropolitan Opportunity: A Reconnaissance and Conceptual Framework.” *Housing Policy Debate* 6 (1): 7–43. https://doi.org/10.1080/10511482.1995.9521183.

Gibbons, Stephen, and Stephen Machin. 2008. *Valuing School Quality, Better Transport, and Lower Crime: Evidence from House Prices.* Oxford Review of Economic Policy 24(1): 99–119. https://doi.org/10.1093/oxrep/grn008

Holvandus, Johanna, and Kadri Leetmaa. 2016. ‘The Views of Neighbourhood Associations on Collaborative Urban Governance in Tallinn, Estonia’. *plaNext - Next Generation Planning* 3:49–66. https://doi.org/10.24306/plnxt.2016.03.004.

Imbens, Guido W., and David S. Lemieux. 2008. *Regression Discontinuity Designs: A Guide to Practice.* Journal of Econometrics 142(2): 615–635. https://doi.org/10.1016/j.jeconom.2007.05.001

Lyons, R. F., and S. Loveridge. 1993. *An Hedonic Estimation of the Effect of Federally Subsidized Housing on Nearby Residential Property Values*. Working paper. St. Paul, MN: Department of Agricultural and Applied Economics, University of Minnesota.

Malone, Thom. 2020. 'There Goes the Neighborhood Does Tipping Exist amongst Income Groups?' *Journal of Housing Economics* 48 (June):101667. https://doi.org/10.1016/j.jhe.2019.101667.

Malpezzi, Stephen. 2008. "Hedonic Pricing Models: A Selective and Applied Review." In *Housing Economics and Public Policy*, edited by Tony O'Sullivan and Kenneth Gibb, 67–89. Oxford: Blackwell Science.

Municipality of Amsterdam. 2025. *Wonen – Stedelijke Ontwikkeling*. Accessed June 16, 2025. https://www.amsterdam.nl/bestuur-organisatie/volg-beleid.

Neumann, Uwe, and Serife Yasar. 2024. "Fostering Prosperity at the Local Scale: Outcomes of Urban Policy for Deprived Neighbourhoods in Germany." *Ruhr Economic Papers* 1129. Essen: RWI – Leibniz-Institut für Wirtschaftsforschung. https://doi.org/10.4419/96973311.

Nguyen, Mai Thi. 2005. 'Does Affordable Housing Detrimentally Affect Property Values? A Review of the Literature'. *Journal of Planning Literature* 20 (1): 15–26. https://doi.org/10.1177/0885412205277069.

OECD. 2021. *Income Inequality and Poverty by Country*. In Income Distribution Database. Paris: OECD Publishing. https://www.oecd.org/social/income-distribution-database.htm

Openshaw, Stan. 1984. *The Modifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography 38. Norwich: Geo Books.

Permentier, Matthieu, Maarten Van Ham, and Gideon Bolt. 2009. 'Neighbourhood Reputation and the Intention to Leave the Neighbourhood'. *Environment and Planning A: Economy and Space* 41 (9): 2162–80. https://doi.org/10.1068/a41262.

Rosen, Sherwin. 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.

Rosenthal, Stuart S., and Stephen L. Ross. 2015. ‘Change and Persistence in the Economic Status of Neighborhoods and Cities’. In *Handbook of Regional and Urban Economics*, 5:1047–1120. Elsevier. https://doi.org/10.1016/B978-0-444-59531-7.00016-8.

Saiz, Albert, and Susan Wachter. 2011. "Immigration and the Neighborhood." *American Economic Journal: Economic Policy* 3 (2): 169–88. https://doi.org/10.1257/pol.3.2.169.

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1 (2): 143–86. https://doi.org/10.1080/0022250X.1971.9989794.

Sirgy, M. Joseph, and T. Cornwell. 2002. “How Neighborhood Features Affect Quality of Life.” *Social Indicators Research* 59 (1): 79–114. https://doi.org/10.1023/A:1016021108513.

Sirmans, Stacy G., David A. Macpherson, and Emily N. Zietz. 2005. "The Composition of Hedonic Pricing Models." *Journal of Real Estate Literature* 13 (1): 1–44. https://doi.org/10.1080/10835547.2005.12090154.

Tomal, Jabed H., and Hafizur Rahman. 2021. “A Bayesian Piecewise Linear Model for the Detection of Breakpoints in Housing Prices.” *Metron* 79 (3): 361–81. https://doi.org/10.1007/s40300-021-00223-8.

Van Gent, Wouter, and Cody Hochstenbach. 2019. ‘The Neo-Liberal Politics and Socio-Spatial Implications of Dutch Post-Crisis Social Housing Policies’. *International Journal of Housing Policy* 20 (1): 156–72. https://doi.org/10.1080/19491247.2019.1682234.

Van Ham, Maarten, David Manley, and Nick Bailey. 2012. "Neighbourhood Effects Research: New Perspectives." In *Neighbourhood Effects Research: New Perspectives*, edited by Maarten van Ham, David Manley, Nick Bailey, Ludi Simpson, and Duncan Maclennan, 1–24. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-2309-2_1.

Whitehead, Christine M. E., ed. 2007. *Social Housing in Europe*. London: London School of Economics and Political Science.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Boca Raton, FL: CRC Press.

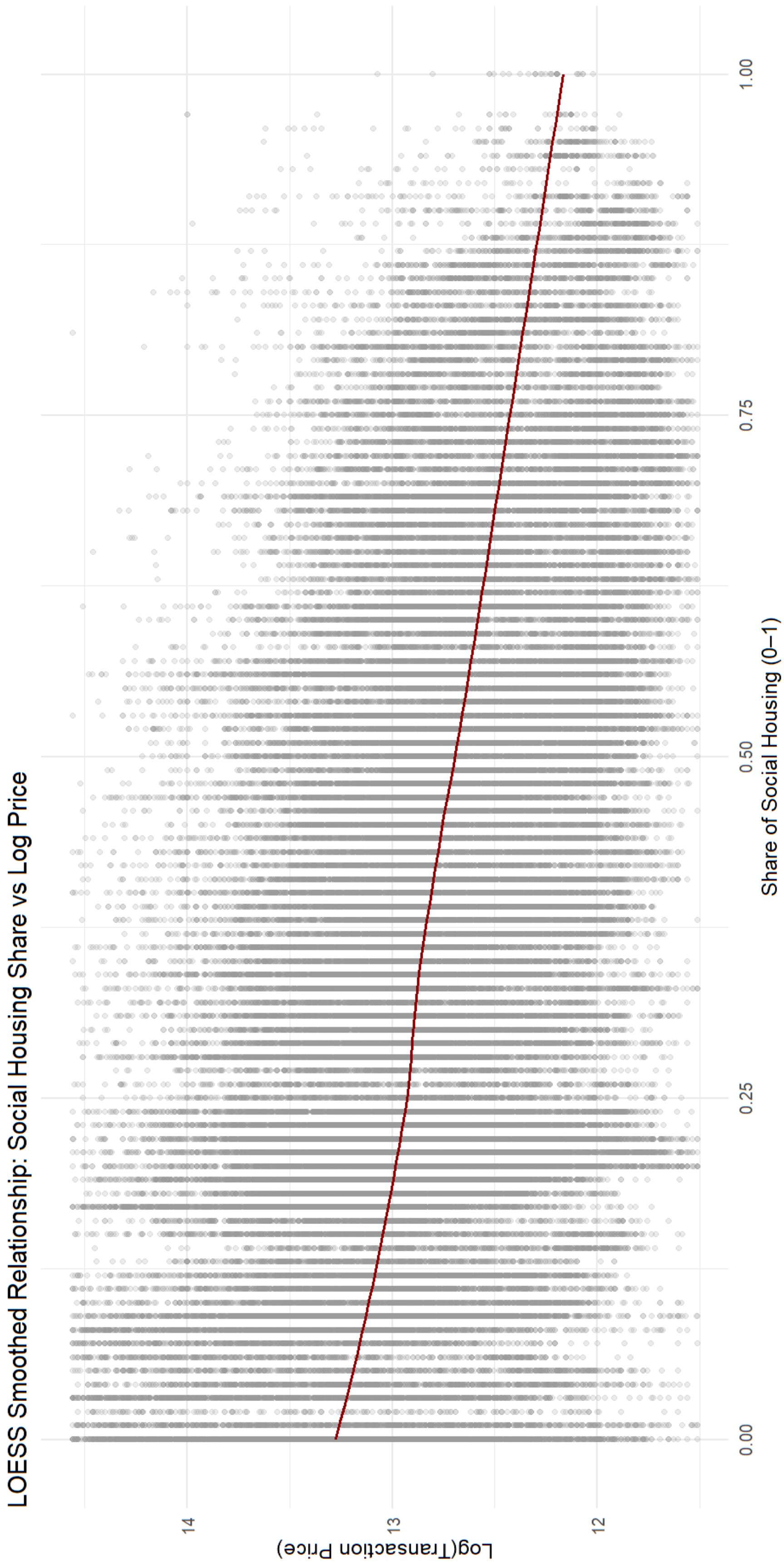
Wooldridge, Jeffrey M. 2021. 'Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3906345.

Zorlu, Aslan, and Jan Latten. 2009. 'Ethnic Sorting in The Netherlands'. *Urban Studies* 46 (9): 1899–1923. https://doi.org/10.1177/0042098009106023.

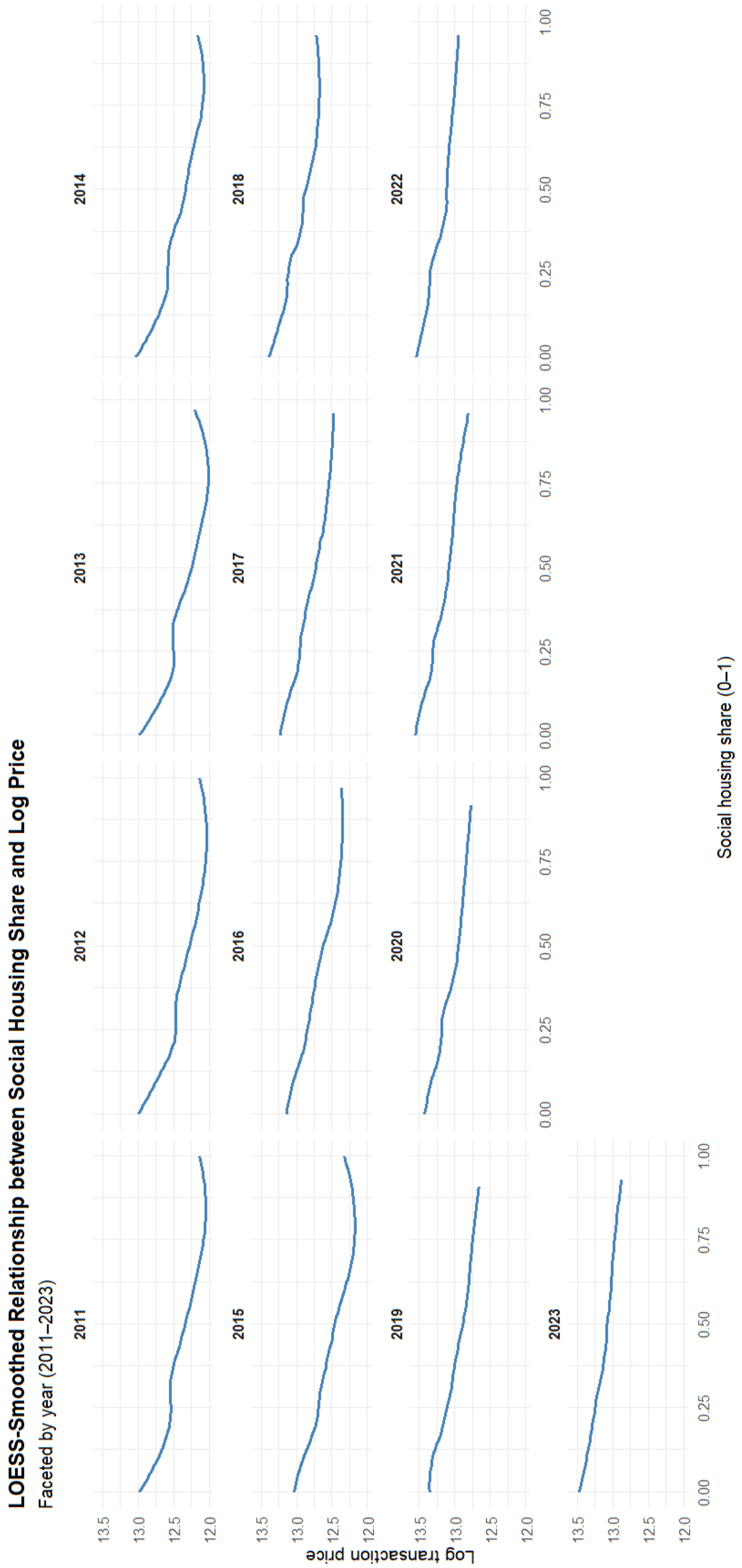
9. Datasets, supportive files and appendices

File name	File type
Dataset 1 – full NVM dataset	CSV
Dataset 2 – neighbourhood level data social housing manually extracted	CSV
Dataset 3 – cleaned NVM dataset Amsterdam	CSV
Dataset 4 – CBS neighbourhood codes	Excel
Dataset 5 – share of social housing file with correct neighbourhood codes	CSV
Dataset 6 – merged dataset NVM and social housing share	CSV
Supportive file 1 – full R script (script 1 – 14)	R
Supportive file 2 – Dashboard Kerncijfers	PDF
Supportive file 3 – Final feedback plan	PDF
Supportive file 4 – Statement on AI use	PDF
R-script 1 – data cleaning and merging	R
R-script 2 – loess plot	R
R-script 3 – binned scatterplot	R
R-script 4 – faceted loess plot	R
R-script 5 – baseline gam	R
R-script 6 – baseline plm	R
R-script 7 – focused grid search plm	R
R-script 8 – neighbourhood fe plm	R
R-script 9 – lagged models gam	R
R-script 10 – lagged models plm	R
R-script 11 – mundlak	R
R-script 12 – hedonic	R
R-script 13 – rdrobust	R
R-script 14 – descriptive statistics	R

Appendix A – Default LOESS plot



Appendix B – Faceted LOESS 2011 – 2023 plots



Appendix C – GAM Control Variables and Significance tables

Variable	Coefficient [95% confidence interval]
Living Area (per 10 m^2)	0.001 [0.000, 0.001]
Number of Rooms	0.187 [0.185, 0.188]
Dwelling Type	0.013 [0.012, 0.013]
Construction Period	-0.017 [-0.017, -0.016]
Num.Obs.	104,716
R2	0.644

Term	edf	Ref.df	F	p-value
s(social_housing_share)	8.91	8.997	2045.64	<0.001

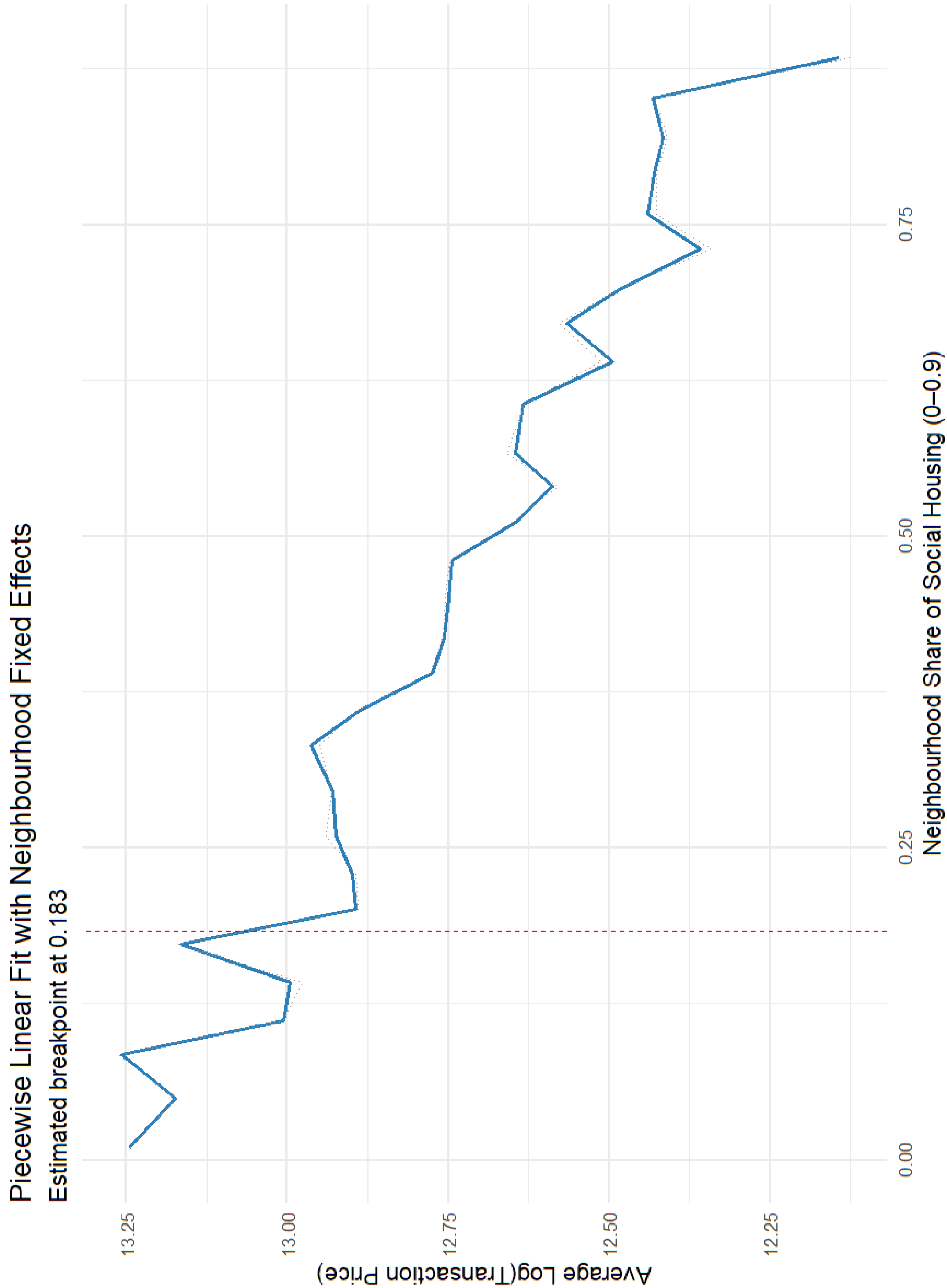
Appendix D – Baseline Algorithmic PLM table

	Estimate	Std. Error	t-value	p-value
Slope below breakpoint	-0.754	0.0371	-20.34	<0.001
Change in slope (U1.S_nt)	0.097	0.0377	2.57	0.010
Estimated breakpoint (psi)	0.190	0.0505		

Appendix E – Focused Grid Search PLM table

Statistic	Value
Estimated breakpoint	0.145
Adjusted R^2	0.643
RMSE	0.350
Slope below breakpoint	-0.833
Slope change above breakpoint	0.192
Slope above breakpoint	-0.641

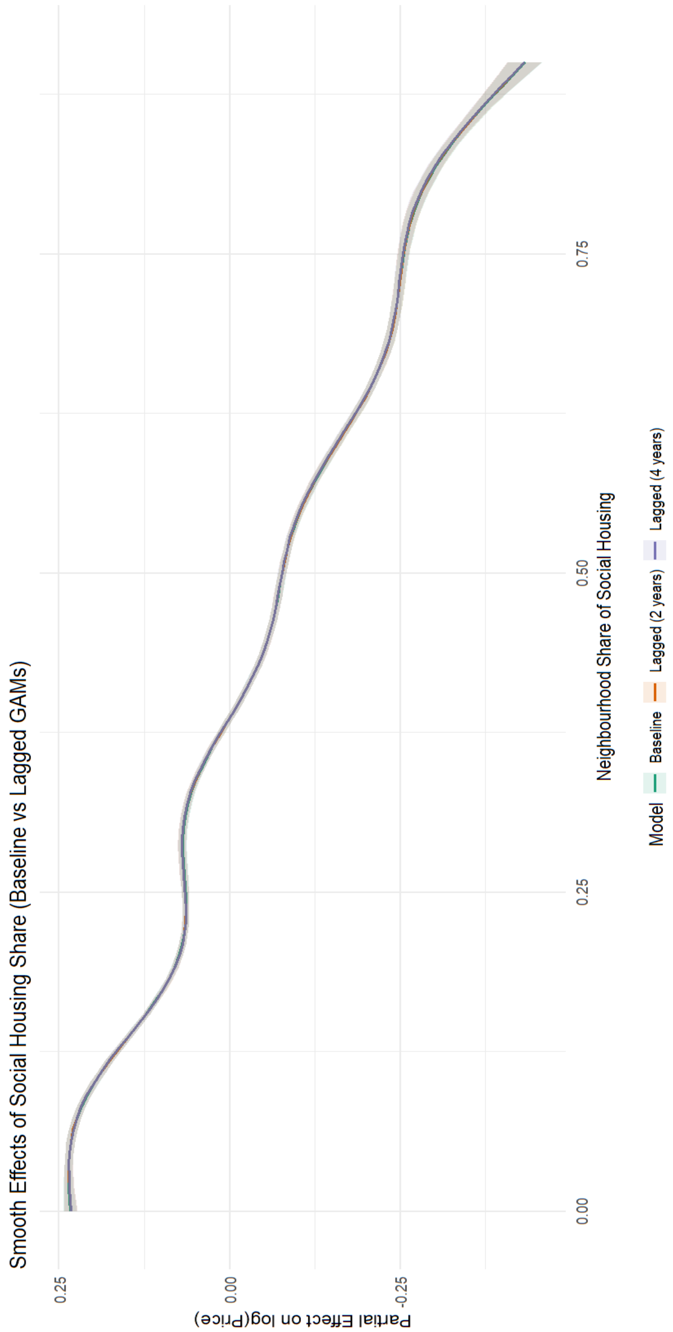
Appendix F – PLM with Neighbourhood FE plot



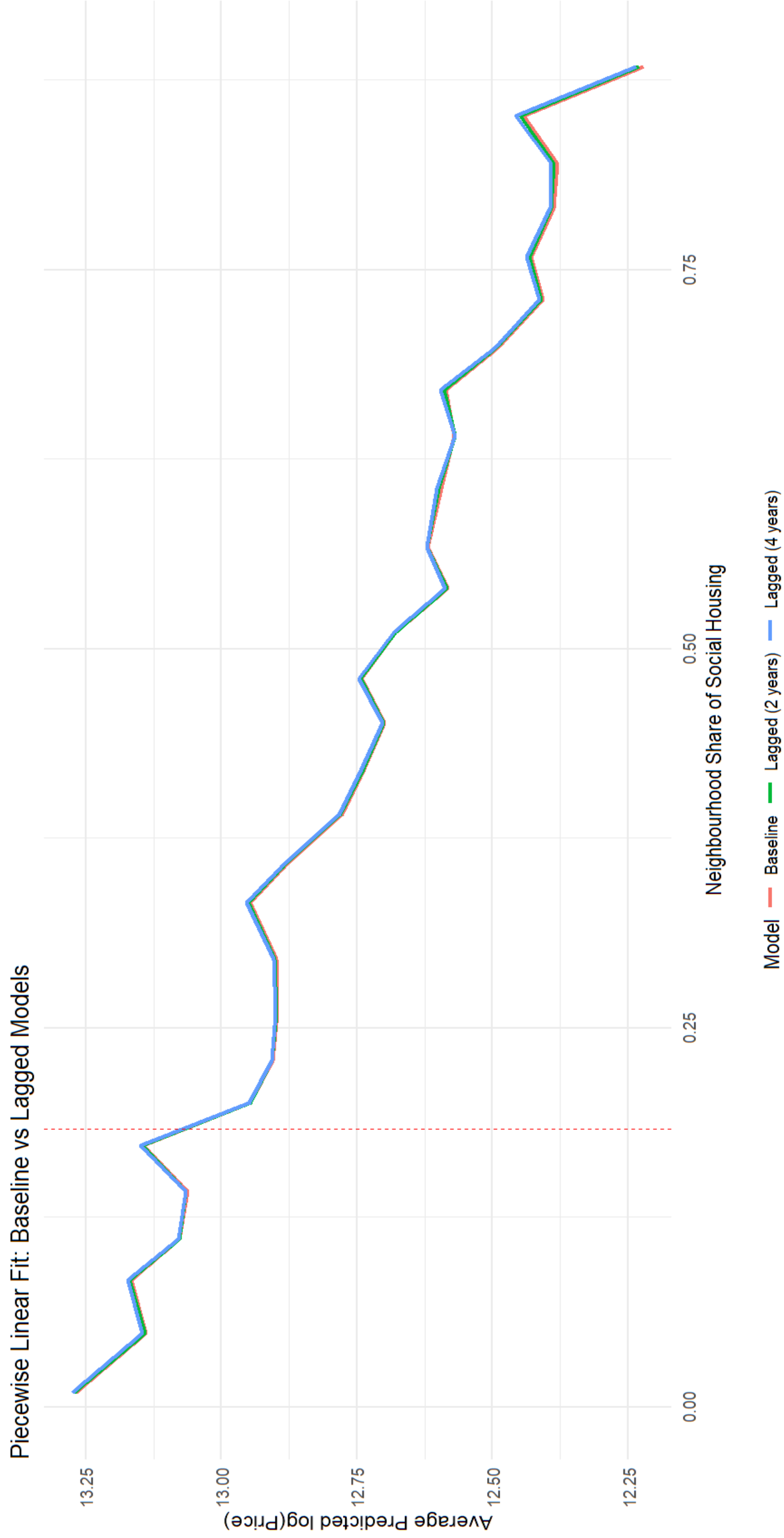
Appendix G – PLM with Neighbourhood FE table

	Estimate	Std. Error	t-value	p-value
Slope below breakpoint	0.183	0.111	1.651	0.0988
Change in slope (U1.S_nt)	-0.758	0.115	-6.600	4.14e-11
Estimated breakpoint (psi)	0.183	0.009		

Appendix H – GAM Lagged Models plot



Appendix I – PLM Lagged Models plot



Appendix J – Model fit statistics Lagged Models table

Model	Adjusted R^2	GCV/UBRE	Estimated Breakpoint
Baseline GAM	0.6430	0.1219	N/a
Lagged GAM (2 years)	0.6431	0.1217	N/a
Lagged GAM (4 years)	0.6431	0.1215	N/a
Baseline PLM	0.642	0.350	0.190
Lagged PLM (2 years)	0.642	0.349	0.190
Lagged PLM (4 years)	0.642	0.349	0.190

Appendix K – Mundlak model table

Variable	Coefficient	Std. Error	p-value
Intercept	12.0800	0.0057	<0.001
Social housing share (within)	-0.6090	0.0353	<0.001
Mean social housing share (between)	-0.0510	0.0353	0.150
Living area (per 10m^2)	0.0006	0.0000	<0.001
Number of rooms	0.1870	0.0008	<0.001
Dwelling type	0.0127	0.0002	<0.001
Construction period	-0.0163	0.0004	<0.001

Appendix L – Summary statistics for living area

Statistic	Value
Mean living area (m^2)	85.87
Median living area (m^2)	76.00
1st Quartile	57.00
3rd Quartile	103.00
Min. living area	10
Max. living area	500

Appendix M – Summary statistics for number of rooms

Statistic	Value
Mean number of rooms	3.38
Median number of rooms	3.00
1st Quartile	3.00
3rd Quartile	4.00
Min. living area	1
Max. living area	10

Appendix N – Distribution of dwelling types

Dwelling type	n	Percentage
Apartment	91,915	87.8
Terraced / semi-detached	11,799	11.3
Detached / other	1,002	1.0