SCHOOL OF
ECONOMICS AND
MANAGEMENT

LUND UNIVERSITY

Master's Programme in Innovation and Global Sustainable Development

Access to Green Spaces and Equitable Economic Growth

The impact of Copenhagen's green infrastructure on socio-economic mobility

by

Mads Stub Wang Pullich
ma5011pu-s@student.lu.se

The aim of Copenhagen municipality to attract socio-economically strong residents with its green infrastructure strategy inspired the analysis which aimed to assess whether improving access green areas in the city has had any impact on socio-economic mobility of its residents. Additionally, the concerns of potential displacement of socio-economically weaker residents warranted consideration of how income inequality affects the impact of access to green spaces. The objective of the research thus to investigate how access to green spaces within 300 meters is related to socioeconomic migration in Copenhagen, specifically if the development of access to green spaces in Copenhagen municipality significantly attracted higher income residents, and if this effect have simultaneously displaced low-income residents. The scope of the analysis covers the 67 neighborhoods in Copenhagen over the year intervals of 2008, 2012 and 2018. The research supported the hypothesis that overall increased access to green spaces has a positive impact on attracting increasing income levels through migration, and especially so at negative migration levels, which suggested that displacement of poorer residents occurs. Furthermore, it specified how the effect of access to green space varies for areas with different income inequality levels, finding that with increasing levels of access to green space in a neighborhood the gentrification effect is mostly evident in areas with high income inequality, while areas with lower income the effects of green space show less evidence of gentrification. The findings should be of interest to historic as well as current green infrastructure policies of Copenhagen municipality aiming toward improvements of agglomeration economics through urban environmental sustainability, while having to balance these changes with measures to ensure socially responsible outcomes. The results of the analysis may help narrow down the parameters and conditions where green infrastructure policies can achieve their desired effects.

EKHS35
Master's Thesis (15 credits ECTS)
August 2022
Supervisor: Kerstin Enflo
Examiner: Jeanne Cilliers
Word Count: 16938

Acknowledgements

This thesis is dedicated to my beloved wife, who’s support has been invaluable and without which the thesis would have never been finished. It is also dedicated to my baby daughter, without whom this thesis would likely have been completed a lot sooner.

Table of Contents

List of Tables

List of Figures

1 Introduction

Urban areas are the source of many of today’s environmental challenges – not surprisingly, since two out of three Europeans live in towns and cities. To this end, local governments and authorities can provide the commitment and innovation needed to tackle and resolve many of these problems (European Commision, 2022). Green space as an urban public good has distinct characteristics that benefit health and recreation, reduce flash floods and improve air quality and urban climate (Kim & Wu, 2022). Green areas thus play a key role in improving urban quality of life and sustainability, and thus green areas should be carefully accounted and evaluated in the urban planning (Cömertler, 2017). Hence, the European Commission's European Green Capital Award recognizes and rewards local efforts to improve the environment, and thereby the economy and the quality of life in cities. The award is given each year to a city, which lead the way in environmentally friendly urban living, and encourages cities to commit to ambitious goals for environmental improvement, attaching great importance to urban green areas (European Commision, 2022). Scandinavian countries are often mentioned as forerunners in sustainable urban development. Green infrastructure planning has played an important role and in practice largely takes the place of the ecosystem services, where social values are the main focus of the plans, particularly in regards to recreation and access to green (Nordh & Olafsson, 2021). As part of the green infrastructure strategy the development of urban green spaces can alleviate the sustainability challenges through benefits to biodiversity and climate change adaptation, as well as recreational impacts on life quality and health. Local policy makers are increasingly recognizing the potential benefits associated with the reduction of distances to green space (Schipperijn et al., 2010), and Copenhagen has become a leading city in sustainable and environmentally friendly urban green places with potential as a role model for both small and big cities around the world. In 2014, Copenhagen was awarded the European Green Capital as a highly successful model for the rest of Europe (Cömertler, 2017).

In Copenhagen access to green areas is a seemingly major factor in attraction of urban residents (Nielsen & Hansen, 2006), as better access to green spaces can benefit residents in the area and create activity opportunities, which may help attract more affluent residents, as was the case in the Vesterbro district of Copenhagen. Urban green areas have been shown to help make the city a great place to live by improving climate and life quality and are thus a key priority when people choose a home in Copenhagen. Good accessibility to green spaces is increasingly becoming a parameter in the competition between metropolitan areas to attract and retain investment, businesses, research, and talents. Livability is thus a crucial position of strength for Copenhagen, and as such it is essential that the capital's position as active, attractive, and green is maintained until 2030, even under population growth pressure. The real estate market consistently demonstrates that many people are willing to pay higher prices for a property located close to a green space. As such, proximity to green areas has a positive influence on property values and a significant impact on urban life in the local area.

The higher values of these residences result in higher property taxes, which makes the process an interest to policymaking. Therefore, the City of Copenhagen considers improving access to green spaces as an important resource for the city's economic development and aims to increase easy and free access to green spaces to attract citizens and create growth. While enjoying a high standard of green accessibility in an international perspective, Copenhagen aims to occupy a third-place in accessibility to green spaces, surpassed only by Stockholm and Helsinki (Government of Denmark, 2019; The Technical and Environmental Administration, 2015). The goal is thus to make Copenhagen a greener city, with 75% of Copenhageners experiencing Copenhagen as a green city by 2025 (The Technical and Environmental Administration, 2018).

The municipality has long worked deliberately with access to green spaces as a factor in retaining and attracting citizens to the municipality to create growth (Rosenbak & Jørgensen, 2009). Since the 1990s Copenhagen been transformed from a challenged metropolis into an attractive alternative to the suburbs for the many who want good housing, green communities, short commuting distances and a safe environment combined with a vibrant and diverse urban life, self-realization, and career opportunities. As such, Copenhagen attracts and retains to a greater extent the able-bodied and resourceful population groups that previously moved out of the city, and the demand for housing is increasing – both for internal migration and among newcomers. This development is seen as combinedly resulting from housing policy, new outlays for residential purposes, and massive public investments in the metro, urban environment, and cultural and leisure facilities. While net immigration numbers varied in the post-1995 period, there has since 2007 steadily been a positive net immigration. However, 52 percent of the population still report that short distance to green areas as a priority when choosing where to live in Copenhagen, mainly consisting of the population segments of families with children and adults without children, while scoring lower in priority for elderly and lowest for the young. In order to continue to attract and retain these segments, it is therefore still seen as important that Copenhagen continues to offer a housing market with attractive nearby green spaces (Center for Urban Development, 2014).

Hence in Copenhagen, green infrastructure is promoted in local strategies, and Copenhagen has worked to ensure that the city's urban green space is developed and maintained to enhance biodiversity and nature experiences, as well was with concern for cultural history, recreational needs, and biological considerations, and actively supports green initiatives on non-municipal land by motivating and engaging in private partnerships (Heward, 2022; Koefoed, 2019). Copenhagen has over the past decade adopted several strategies which are relevant to the creation and support of green infrastructure, and has developed a number of strategies to ensure involvement of the city's green structure in other municipal planning, and the quality of parks and natural areas is being maintained and developed (Cömertler, 2017), including:

In 2007 Copenhagen adopted the policy "Eco-metropolis - Our vision for Copenhagen 2015," which aims for the city to become a "green and blue capital city," where reduction of distance to urban green areas is an overall goal for blue and green area development (Cömertler, 2017).

The strategy "Nature in Copenhagen 2015-2025" aims to ensure that the city develops into a "green and climate-friendly" city, with the primary goals to create more nature in Copenhagen and to improve the quality of the natural areas in Copenhagen.

Additionally, the "CPH 2025 Climate Plan" running from 2011 - 2025, the city aims to become Europe's first carbon neutral city by 2025, which includes the creation of additional larger green areas as well as smaller greening throughout the city.

Finally, "Co-create Copenhagen" is a vision targeting technical and environmental issues towards 2025, and aims for Copenhagen to be "a livable city," "a city with an edge" and "a responsible city". One of the tools for achieving this vision has been to enhance access to urban green spaces (Green Infrastructure, 2022).

However, promoting green space access does not always provide socioeconomic benefits. Gentrification and displacement of marginalized long-time residents resulting from rising property values in response to new and restored green spaces is a especially of critical concern (Goossens, Oosterlynck & Bradt, 2020). Mobility patterns show the movement of low-income people from high-green areas to low-green areas over time, influencing the replacement of the poor with wealthier inhabitants, and worsening inequality in green access (Sharifi et al., 2021). Urban green spaces may thus have positive and negative immediate impacts on the residential well-being, residential location choice, housing, and land markets. Policy-makers need to know how access to public green space varies across society, and whether those who enjoy the greatest access include those who are most in need, and should thus consider both positive and negative factors in the development of green spaces. (Schwarz et al., 2021).

The City of Copenhagen thus aims to improve access to green areas high quality in Copenhagen, with the municipal plan 2015 stating that the development of the urban green areas must be seen in the context of the development and composition of the population and that the City of Copenhagen must promote the establishment of publicly accessible green areas as part of the urban development. In order to ensure effective and equitable effects of improved access to urban green space, a thorough analysis of local neighborhoods, their population, and the available green spaces is needed before deciding on a viable strategy to increase the use of green space (The Technical and Environmental Administration, 2015).

1.1 Aim and Scope

Considering the aim of Copenhagen municipality to attract socio-economically strong residents with its green infrastructure strategy it would be interesting to investigate whether improving access green areas in the city has had any impact on socio-economic mobility of its residents. Additionally, the concerns of potential displacement of socio-economically weaker residents warrants a consideration of how income inequality affects the impact of access to green spaces. As such, the aim of this paper is to investigate the following research question:

How is access to green spaces in the city of Copenhagen related to socioeconomic migration? Specifically, has the development of access to green spaces in Copenhagen municipality significantly attracted higher income residents, and/or displaced lower income residents?

The geographical scope of the analysis is limited to the 67 different neighborhoods inside Copenhagen municipality, while the temporal scope of the analysis is restricted by availability of data on green spaces in the study area and covers the interval of years 2008, 2012 and 2018.

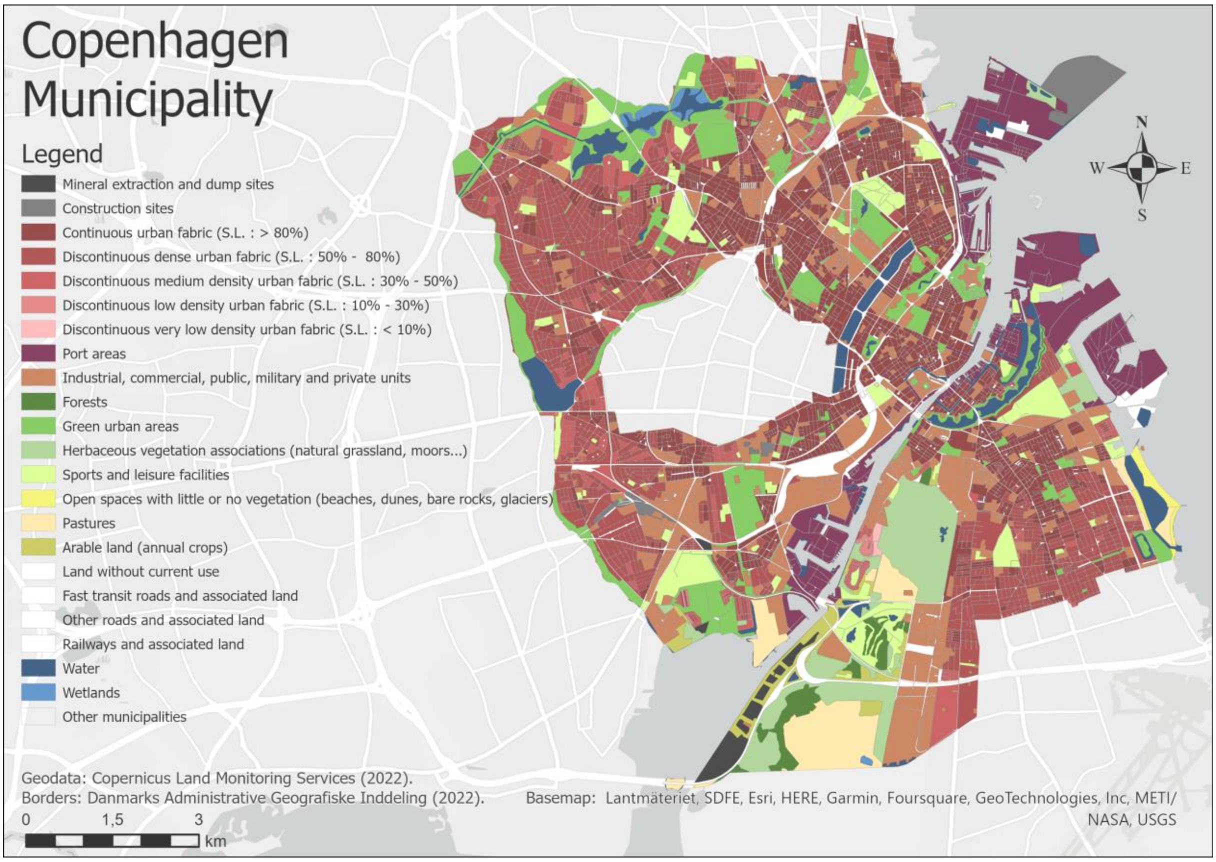


Figure 1. Reference map of Copenhagen municipality (2018). Cartography by the author.

1.2 Outline of the Thesis

Chapter 1 covers the introduction to the green infrastructure policies of Copenhagen, where increasing access to green space is a key to its environmental and socioeconomic objectives. Chapter 2 investigates the literature review of theory on the effects of access to green space on socioeconomic migration, from which the hypothesized interaction models are derived. Chapter 3 overviews the key variables of wage, net immigration, green service area, gini index, as well as control variables, and considers the ecological inference of aggregated panel data. Chapter 4 outlines the methodology used to conduct the analysis, namely analysis of margins in fixed effects regressions with two-way and three-way interactions between the key variables. Chapter 5 examines and the results of the average marginal effects and predictive margins, and discusses the validity of the findings, which ultimately support the hypothesized relationships.

2 Theory

This chapter examines the extant literature of research on the topics of the effects related to access to urban green spaces relating to its effects on socioeconomic mobility of residents in general and in Copenhagen in particular, notably regarding hedonic pricing and gentrification. Additionally, research on the potential parameters of these effects are covered, as well as the aims and goals of the green infrastructure policies of Copenhagen municipality in this regard.

2.1 Previous Research

Service area of 300 meters to green spaces

Several Danish studies have found a clear correlation between distance to green areas and usage (Holm & Jakobsen, 2006; Nielsen & Hansen, 2006; Rosenbak & Jørgensen, 2009). Research from the Netherlands similarly documented this correlation between distance and usage regardless of degree of urbanity, size or shape, and underline the importance of accessibility (Bijker & Sijtsma, 2017; Maas et al., 2006). While the European Environment Agency (EEA) recommends that people should have access to green space within 15 min walking distance (roughly 900 meters), public agencies of other countries such as the English Nature in England recommend that people should live no further than 300 meters from the nearest green space (Barbosa et al., 2007). More than 80 percent of citizens in most of the Green Capitals live within 300 meters of a green area, including Copenhagen when counting small green spaces (Cömertler, 2017), while Danish public estimates find less than 50 percent of Copenhageners living within 300 meters of green space when considering only medium and large green areas (The Technical and Environmental Administration, 2018). In Denmark, living within 300 meters from a green space has also been found to be the distance parameter where a significant increase in attraction, attendance, and health benefits occurs (Stigsdotter et al., 2010). Public assessments also found a correlation between distance, usage, and health at around 300-400 meters while investigating the need for new large and medium-sized green spaces in Copenhagen at the service goals 500 meters, 800 meters, and 1000 meters. The study found no clear answer to which service target should be set, but concluded that the existing service target of 1000 meters did not correspond to the recommendations of the research results, and was unambitious compared to other cities, especially the Green Capitals (The Technical and Environmental Administration, 2018). The government has since asserted a vision for every Copenhagener to live within 300 meters of a green space (Politiken, 2019). Analysis of neighborhoods, population, and the availability of green spaces is recommended before deciding on a viable strategy to increase green space in the city (Schipperijn et al., 2010).

Hedonic pricing, residential mobility, and green spaces

Studies have indeed found a that access to nearby green space is attractive in the housing market (Evangelio et al., 2019; Gehl, Kaefer & Reigstad, 2006; Kong, Yin & Nakagoshi, 2007; Luttik, 2000; Razykova, Bartke & Schwarz, 2014; Sullivan, Kuo & Depooter, 2004). Specifically, households respond to variation in green space with their housing location decisions to the extent that the available properties and their income and lifestyle constraints permit (Cheshire & Sheppard, 2005; Sheppard & Cheshire, 1995; Wu & Plantinga, 2003). The variety of urban green spaces, i.e. trees, urban parks, forests, and backyards provide urban ecosystem services: recreation, local climate regulation, as well as air quality improvement that might considerably influence where people prefer to live. **Residential mobility** and locational choice are as such major drivers of urban land use change (Razykova, Bartke & Schwarz, 2014).

The real estate market consistently demonstrates that many people are willing to pay higher **prices for a property located close to a green space**. Additionally, the higher values of these residences result in higher property taxes, which makes the process an interest to policymaking. This process of capitalization of green spaces into the value of nearby properties is termed the "proximate principle, whereby the competitive market conceptually will bid up property values equal to the capitalized value of the benefits that property owners perceive they receive from the presence of green spaces, a mechanism known as "hedonic pricing". A comprehensive literature study found that the capitalization of benefits ceased at selected distances, usually between 150 meters and 1000 meters from green spaces in urban contexts (Crompton, 2001).

Hedonic pricing based on the distance to green spaces seemingly also occurs in Copenhagen. The project "Gains from investments in urban life and urban quality" (a collaboration between some Danish municipalities, City and Port Development, and University of Copenhagen) concludes that there is a direct correlation between proximity to green areas and property value. Overall in Danish cities the value of a property increases by up to 10% on average for every additional 10 hectares of park or urban natural area found within 500 meters walking distance, while both valuation and range are slightly lower for apartments in cities such as Copenhagen (The Technical and Environmental Administration, 2018).

Gentrification and green spaces

In the context of hedonic pricing and residential mobility, promoting green space access does not always provide socioeconomic benefits. Gentrification and displacement of marginalized long-time residents resulting from rising property values in response to new and restored green spaces has been documented extensively (Goossens, Oosterlynck & Bradt, 2020; Rigolon & Nemeth, 2018) (Anguelovski et al., 2018; Wolch, Byrne & Newell, 2014) (Connolly, 2019; Haase et al., 2017; Immergluck & Balan, 2018; Rigolon & Németh, 2020). Mobility patterns show the movement of low-income people from high-green areas to low-green areas over time, influencing the replacement of the poor with wealthier inhabitants, and worsening inequality in green access. Yet, this focus is recent, and there is a gap in the knowledge on the role of population mobility and residential relocation in shaping urban spatial patterns over time (Sharifi et al., 2021).Thus, the relationship between urban green spaces and residential development is complex: Urban green spaces have positive and negative immediate impacts on residential well-being, residential location choice, housing, and land markets.

As such, urban planners need consider both positive and negative factors in the development of green spaces (Schwarz et al., 2021). Strong short-term green gentrification effect are observed in medium-sized green spaces. Taking these short-term and local-level gentrification effects of green space characteristics into consideration allows more inclusive development and equitable outcomes (Kim & Wu, 2022). This suggests that analyzing medium and large green spaces, without including smaller areas, is a crucial point for policy making (The Technical and Environmental Administration, 2018).

2.2 Theoretical Approach

Based on the findings from the literature review we hypothesize that hedonic pricing based on distance to green increases the average income level of an urban area through the attraction of higher income residents, i.e. better taxpayers, while simultaneously in areas with higher income inequality this population mobility may foster gentrification by displacing long-time lower income residents, thus worsening socioeconomic differences in distance to green space.

The findings from the literature review suggest a focus on the effects of distance to medium and large green spaces at the neighborhood level with a distance parameter set to 300 meters, which should be investigated over a short-term time span to capture the effects of gentrification. In order to investigate the full hypothesized mechanism of access to green space increasing neighborhood income levels by attracting higher-income residents, and potentially displacing lower-income residents, it is first necessary to investigate whether there is an independent interaction in the first relationship between access to green space, migration, and income levels. Thereafter, the second interaction is investigated by including the effect of income inequality. Hence, the thesis conducts a panel regression analysis of a two-way and three-way interaction.

Specifically, expect a two-way interaction with average income levels as response variable, migration levels as the predictor variable, while access to green space act as moderator variable on the relationship between the predictor variable and response variable, shown in Figure 2.

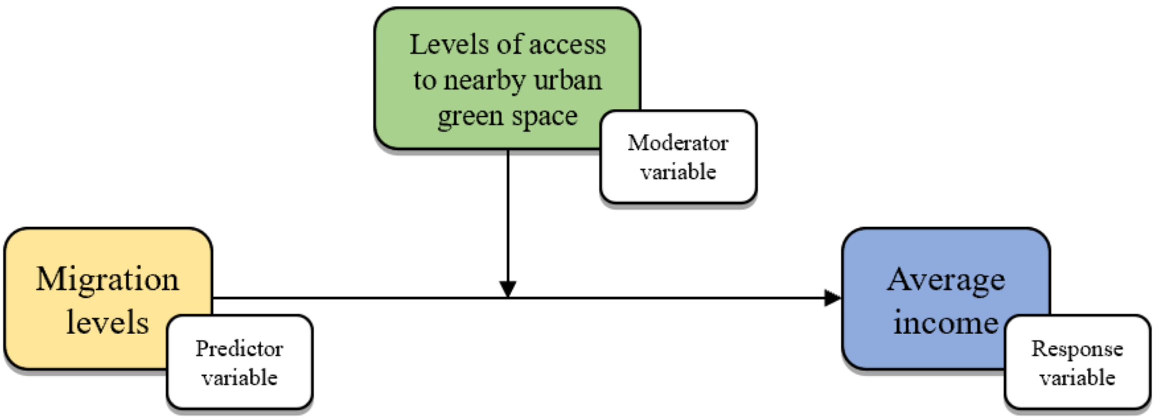


Figure 2. Theoretical model for two-way interaction effects. Author's own work.

If the moderator effect of green space is found to be significant, hence begins an investigation of the expected three-way interaction where average income levels is the response variable, migration is the predictor variable, while both access to green space and income inequality act as moderators for the independent variable, as shown in Figure 3.

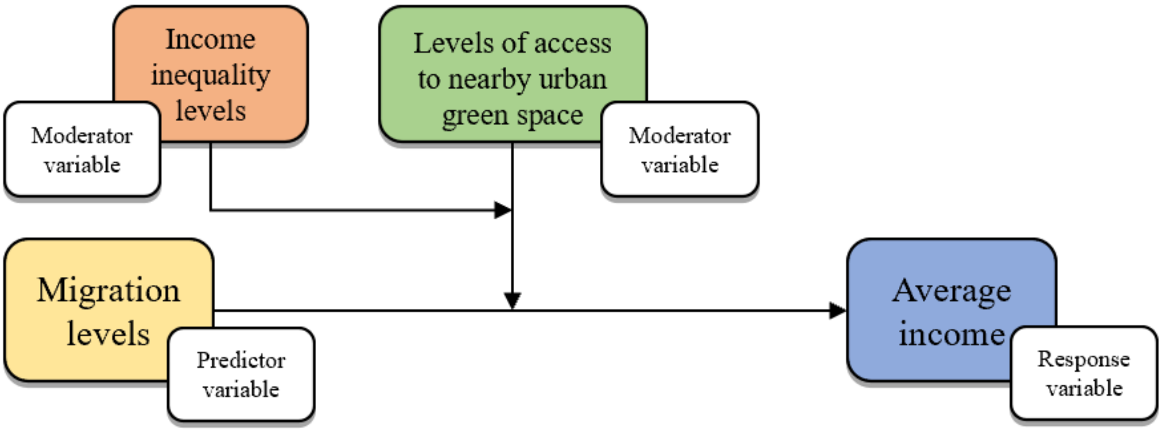


Figure 3. Theoretical model for three-way interaction effects. Author's own work.

The expected key variables are thus "average income" as the response variable, which should be a good direct estimation of the effect of hedonic pricing via access to green space attracting higher income residents to the study area, and potentially displacing lower income residents. We expect the main predictor variable to be "net immigration," while "access to green areas" should be a moderator variable, since there is no expected direct relationship with income, and only an expected effect of green space on income via its interaction with immigration. Also, since the focus in on whether socioeconomic differences in access to green space affect the interaction by displacing lower income residents, thus increasing average income levels, the expectation is that "income inequality" has an effect as a co-moderator with to green access on the relationship between net migration and average income at the neighborhood level.

3 Data

In order to examine whether the relationship between immigration and income levels is moderated by levels of access to green space alone and jointly with income inequality levels secondary quantitative data is collected according to the established theoretical motivations. This chapter covers a critical examination of the sourcing, collection and specification of the different data utilized in the analysis following the guidelines of (Creswell & Creswell, 2018).

3.1 Source Material

Neighborhoods form the panel variable consisting of macrounits of aggregated numbers of residents in each neighborhood and is therefore ecological in nature. Neighborhoods are used as unit of analysis since they are the smallest scale of available socioeconomic statistics for Copenhagen municipality. Neighborhoods are a subdivision of Copenhagen municipality, consisting of 67 areas as shown in Appendix A. The designations of the neighborhoods was created for the purpose of allowing for the smallest geographic division of Copenhagen at which open statistical data can be utilized without potential issues of privacy discretion. Neighborhood designations were designated by the Copenhagen local district committees in 2020, however, official socioeconomic statistics by Statistics Denmark for years earlier than 2020 has been compiled back in time according to neighborhood boundaries (Copenhagen Municipality, 2022). Meanwhile official GIS data from DAGI is used to obtain geographic borders of the neighborhood used to calculate the green service area variable (Agency of Data Supply and Infrastructure, 2022). The small scale of neighborhoods as unit of analysis should allow for a detailed analysis of the effects of access to green space.

Socioeconomic panel data for each neighborhood obtained from Copenhagen Municipality (City of Copenhagen: Statbank - Data and Statistics, 2022) are used to create the response variable Average Income, the predictor variable Net Immigration, one of the two moderator variables Gini Index, as well as theoretically motivated control variables. Thereafter the different extracted statistics are manually combined to achieve the full dataset. The data is collected by Statistics Denmark (StatBank Denmark, 2022), which as the central authority on Danish statistics collects, compiles and publishes statistics on the Danish society. The requirement for official statistics follows on the Act on Statistics Denmark, which provides the formal framework for official statistics in Denmark. The statistics meet the requirements developed in international cooperation, and satisfy quality requirements formulated in the European Statistics European Code of Practice (Quality in the Production of Statistics, 2022). Thus, the data should present the best available quality in reliability and representativity. Validity of the selected data is discussed for each variable in this section of the chapter.

Geographical Information Systems (GIS) data is used to calculate the green space service area. Some considerations must be made when obtaining data for access to green space within 300m, as the data needs to be calculated manually, of which the method will likely dictate the results. Firstly, in order to calculate the 300 meters according to the stated theoretical specifications, geodata consisting of land cover and land use data at a high spatial resolution is needed to avoid omitting medium to large green spaces which could bias the estimates. To this end, geodata for Copenhagen municipality is collected from Copernicus Land Monitoring Service (CLMS), a European programme which collects and combines satellite data with observation data from sensor networks on the earth's surface, which provides geographical information on land cover to a broad range of users in the field of environmental terrestrial applications. CLMS offers two different geodata sets of land cover for Copenhagen: Urban Atlas and CORINE Land Cover. (Urban Atlas — Copernicus Land Monitoring Service, 2022) offers a high spatial resolution with 17 distinct land cover designations for the years 2006, 2012 and 2018. On the other hand, (CORINE Land Cover — Copernicus Land Monitoring Service, 2022) offers geodata for a longer time span of 1990, 2000, 2006, 2012, and 2018 with 44 land use classes, however at a much lower resolution. Upon inspection the CORINE dataset is missing many medium-sized green areas when compared to the data from Urban Atlas. Since focus is on the effect of medium to large green spaces the CORINE dataset is likely too low resolution to provide unbiased estimates. Additionally, the longer time span of CORINE would have been useful for a better estimation of effects over time of green spaces, but some of the key control variables needed for the analysis, such as education, only has data available from year 2008. Thus, the inclusion of years 1990, 2000 in CORINE does not add more utility than Urban Atlas. As such, due to the higher spatial resolution Urban Atlas is deemed the better fit to calculate a more precise and unbiased variable for access to green space.

3.2 Specification & Evaluation

3.2.1 Response Variable: Income

While the theoretical motivation recommends the use of real estate prices as response variable, this data was unavailable. Instead, income serves as the response variable, as it should similarly be able to capture the effects of green space and potentially income inequality via immigration. We use the dataset KKIND3 to obtain statistics on average personal income in 1000 kr. of people above 14 years of age in order to obtain income average only of people for which it is relevant. In year 2008 one neighborhoods has a mean annual income of 0 kr. and a population of 1, probably being either a mistake or an extreme outlier which can cause abnormal distribution and is therefore excluded. As seen in Appendix B, the data is asymmetrically skewed, and does not exhibit a normal Gaussian distribution, which is problematic as it makes realistic interpretation of the regression results difficult. However, incomes commonly follow an exponential distribution, and since our raw income data follows a log-normal distribution, a natural logarithmic transformation is used to specify the distribution in a linear functional form which helps to ease interpretation. The transformed variable, as is visualized in Appendix B, and the raw data are statistically equivalent, and thus the transformed income variable is used, while the main estimates are tested for robustness with estimates derived from the raw variable.

Table 1. Summary statistics of dependent variable "Income" and log transformation

	Obs	Mean	Std. dev.	Min	Max
Average annual income in kr.					
Total	200	316388	82520	182816	635578
2008	66	286081	60232	182816	510834
2012	67	309188	78813	185390	570329
2018	67	353443	91556	200357	635578
(Log)Wage					
Total	198	12,628	0,229	12,116	13,254
2008	66	12,544	0,197	12,116	13,144
2012	67	12,613	0,234	12,130	13,254
2018	65	12,727	0,220	12,208	13,229

3.2.2 Predictor Variable: Net Immigration

Immigration serves as the main independent variable through which the hypothesized effect of the moderator variables access to green service areas and income inequality will be examined. Immigration data is obtained from the dataset KKBEF6 where net immigration numbers are used to capture the effect of both retention, immigration, and emigration, which should be a valid representation of the different facets of the hypothesized effects of the moderators. Examining the summary statistics of the variable in Table 2 it is clear that the change in mean and standard deviation over the time is minute, and as such the variance in the data is largely due to a relatively small number of observations, which is also clear from the visual relationship between wage and net immigration in Figure 4. While this does not necessarily run contrary to the theoretical expectations, it does limit the expected effect of both the moderator variables.

Table 2. Summary statistics of Net Immigration.

Net Immigration	Obs.	Mean	Std. dev.	Min	Max
Total	200	0,011	0,013	-0,035	0,093
2008	66	0,013	0,016	-0,001	0,093
2012	67	0,010	0,009	0,000	0,050
2018	67	0,009	0,013	-0,035	0,052

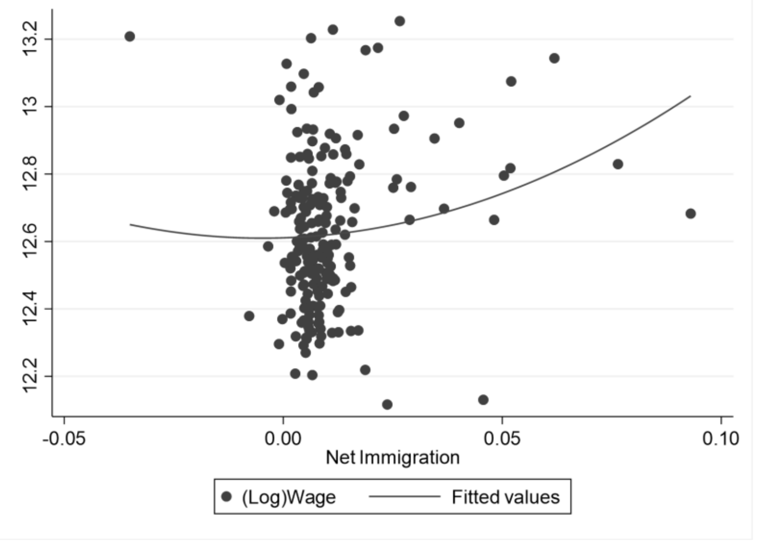


Figure 4. Relationship between (log)Wage and Net Immigration.

3.2.3 Moderator Variable: Income Inequality

Income inequality in each neighborhood is obtained from the dataset KKIND6 as Gini-index, a measure of statistical dispersion used to represent the income inequality within a social group. The Gini index is widely used as a stable measure of income inequality due to its natural characteristics of the Lorenz curve (Farris, 2010; Gastwirth, 1972). The summary statistics in Table 3 how a relatively stable mean and low standard deviation with high variance in minimum and maximum values, and Figure 5 in shows a strong relation between Gini Index and income which follows theoretical expectations, while showing a more minute relationship between Gini Index and Net Immigration, which suggest that the moderator effect is minor.

Table 3. Summary statistics of Gini Index.

	Obs.	Mean	Std. dev.	Min	Max
Gini Index					
Total	200	30,418	5,911	20,48	48,62
2008	66	28,873	5,893	20,48	47,57
2012	67	30,094	5,233	21,22	44,28
2018	67	32,264	6,148	23,59	48,62

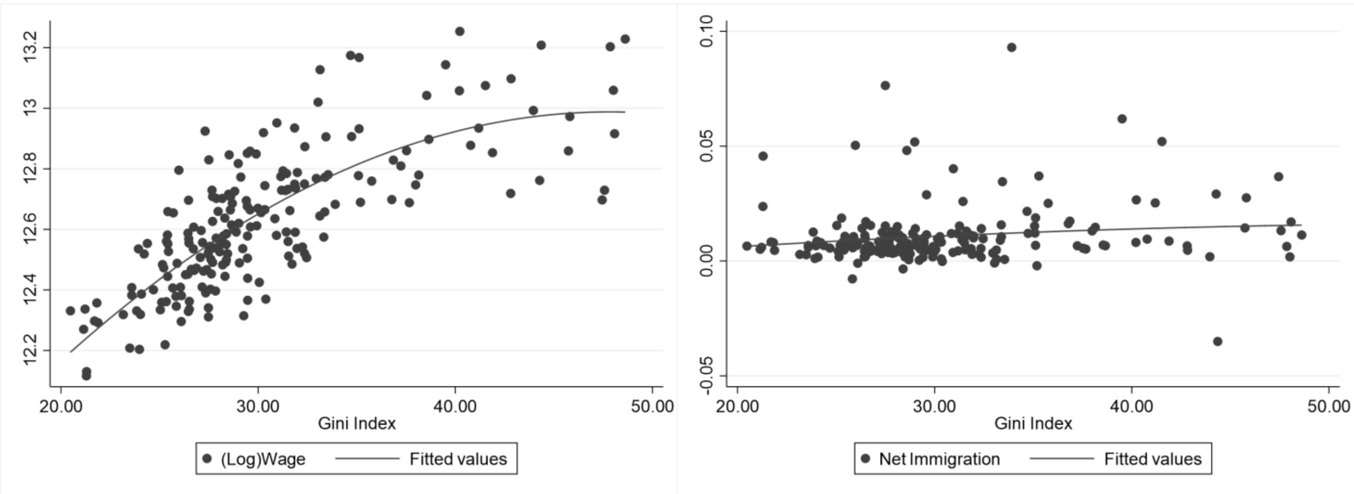


Figure 5. Relationship (log)Wage and Gini Index (left), and Net Immigration and Gini Index (right).

3.2.4 Moderator Variable: Green Service Area

Calculating the access to green areas was done using the GIS software ArcGIS Pro (Esri, 2022). As shown in Appendix C, Urban Atlas 2006 and 2012, 2018 have minor differences in land use classifications, yet these are easily reclassified into useful themes necessary for calculation. Following the same scheme as official research channels (The Technical and Environmental Administration, 2018), the data is classified into: “residential areas”, “green areas”, and other non-relevant areas. While previous studies on access to green space in Copenhagen have utilized Network Analysis using road network and housing geodata to calculate access to green space for each residential building, these methods are unfortunately not possible with the lack of available housing data this study. Network Analysis methods cannot be used in this analysis as there is no publicly available precise population geodata at the time of writing. Instead, a buffer zone of 300 meters around each green area is generated to serve as a proxy service area.

This service area is used to summarize how much of each residential area is covered by the service area, and thus serviced by green spaces within 300 meters. Afterwards the service area are segmented into the geographic borders of each neighborhood. Appendix A shows the product including neighborhood borders, the reclassified green areas, residential / urban areas inside and outside the 300 meter service area, and non-relevant areas. In order to obtain the final variable the green service areas are normalized by the size of the residential area to obtain a percentage and avoid bias due to overall neighborhood size. The service area are not normalized by residential density since green spaces are considered sharable, thus a percentage coverage should be better. The variable should be a somewhat smooth, yet valid proxy, capturing most relevant residents. Importantly, since the control variable for education only goes as far back as 2008 this causes an inconsistency between the green service area variable for year 2006. Since the inclusion of education as a control variable is a necessary to avoid potential issues of omitted variable bias, the green service area for year 2006 is held constant for year 2008 by assuming that the slow changes of the variable allow for a lagged effect.

Table 4. Summary statistics of the created variable "green service area".

	Obs.	Mean	Std. dev.	Min	Max
Green service area					
Total	200	0,531	0,159	0,061	0,841
2008	66	0,527	0,153	0,165	0,831
2012	67	0,533	0,165	0,061	0,826
2018	67	0,533	0,162	0,100	0,841

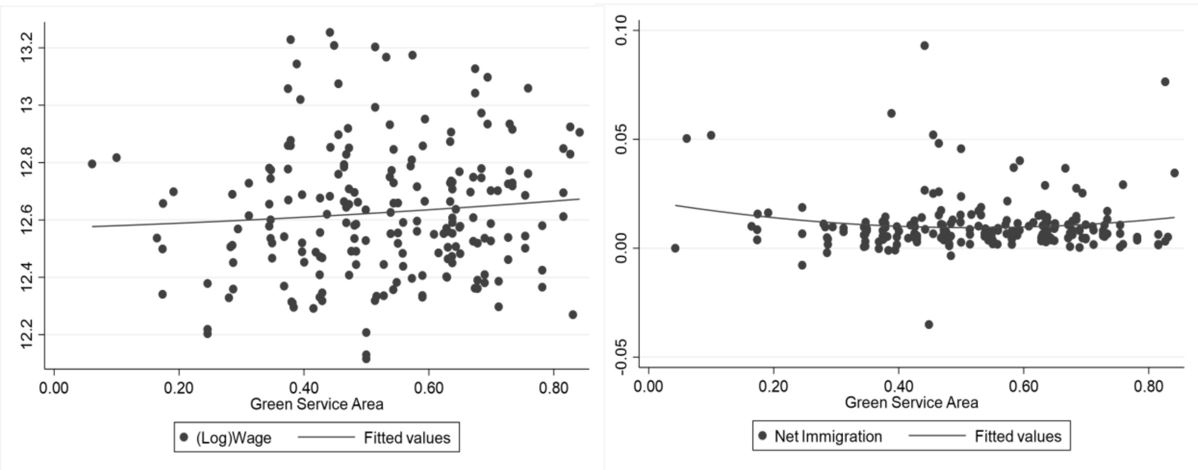


Figure 6. Relation (log)Wage and Service Area (Left), and Net Immigration and Service Area (right).

Summary statistics in Table 4 show very minute changes in the mean service area coverage from year 2008(2006) to 2012, while there is no mean change from 2012 to 2018, which fits our expectations derived from the literature. However, the mean value of 2008 is a little higher than that found in other studies (The Technical and Environmental Administration, 2018), likely due to lack of density in the variable. Yet, the changes in in minimum and maximum coverage values is noticeable, indicating significant variance over time for some neighborhoods. Which may be able to capture effects on immigration and wage. Figure 6 shows no clear relationship between wage and green service area coverage, while showing a minor noticeable relationship between immigration, which follows our expectations, as green spaces should relate to wages only via its effect on migration and retention of productive residents.

3.2.5 Control Variables

Education Levels

The “Mincer earnings equation” is a widely used econometric model in empirical economics, which stipulates that log earnings can be modelled as the sum of a linear function of years of education and as well as a quadratic function of years of potential experience (Mincer, 1974). While modifications have been suggested over the years, it is still considered a robust model (Lemieux, 2006). As such, education should be an important theoretical control variable which should be included in an encompassing model to avoid omitted variable bias. Importantly, data for education is only available starting from year 2008, and since its inclusion in the analysis is needed to avoid potential omitted variable bias, all collected panel data will start at year 2008. The education data is collected from the dataset KKUDD1 which contains categorically aggregated numbers of residents by highest completed education levels. In order to ease the estimation process, these are grouped into categories of similar relations with wages: Elementary school and high school are grouped as “lower education.” Professional school, short upper education and bachelor education are grouped as “middle education.” Finally, both middle and long upper education, and Ph.D. and above are grouped as “upper education.” These variables are then normalized by the total number of residents to obtain a percentage misspecification of variables due to differences in residential numbers in each neighborhood.

Age

In wage regressions a variable for years of education can be problematic if an ability or experience variable is missing from the equation (Kennedy, 2008). At the time of writing there are no available such as aggregated years of working experience, thus aggregated age cohort data is used as a proxy for potential work experience. Potential work experience is usually calculated as age minus years of education minus 6 (school age) to capture the number of years the average person could have been working since graduating (Zveglich Jr., Rodgers & Lavina, 2019). Since the data is ecological, this method would be meaningless, and a different method is used. The dataset KKBEF1 is used to obtain data containing number of residents by 10 year cohorts, which are recoded into three themes: Cohorts 0-19 years are grouped into “young age” which are not expected to have much if any working experience, or income. Thereafter cohorts 70-99+ years are grouped into “retirement age”, which should similarly have lower wages. Cohorts 20-69 years are finally grouped into “productive age”, which includes the productive population with working experience and thus higher incomes (Skirbekk, 2004).

Apart from the modified Mincer equation, additional control variables are included to add explanatory power, including gender ratio, employment, ethnicity, population density, housing. These are discussed further in Appendix I.

3.2.6 Panel Data

With the collected and specified panel data there is thus a strongly balanced panel data set, meaning that all cross-section units have observations on all time periods, which gives us a relatively large N=67 and a small T=3. Ecological socioeconomic panel data such as ours tends to be short and wide, consisting of a very large number of cross-sectional units observed over a small number of time periods, in other words with large N and small T. Using panel data for our analysis has several advantages: It can deal with heterogeneity cross-sectional data by avoiding the omission of time-series variables which may otherwise causes bias in estimation. Similarly it corrects for problems of omitted time series variables that influence the behavior of the units uniformly, but differently in each time period. Additionally, using panel data creates more informative data variability by combining unit variation with time variation, which alleviates multicollinearity problems and allows for more efficient estimation. Additionally, panel data tends to be aggregated, which averages away heterogeneity and leads to data and relationships with continuity and smoothness features. However, statistical results using aggregated data in this context do not necessarily reflect the underlying individual behavioral relationships. Therefore, ecological inference using of aggregate data to study the behavior of individuals (a case such as ours inferring the aggregate migration choices of individuals in relation to their personal preferences of access to green space and the socioeconomic makeup of the area) should be done with great care (Kennedy, 2008).

Ecological inference usually generates inaccurate conclusions about the empirical world, which gives rise to the ecological inference problem. It is, however, necessary for conducting policy analysis where individual-level surveys are unavailable and must therefore be dealt with. The inference problem concerns mainly aggregation bias, which causes information loss when individual-level data are aggregated into the observed marginals. For aggregate data collections sometimes the types of information loss may be selective, which the inference must consider in order to avoid bias. Also, a secondary cause of inaccurate ecological inferences is a variety of basic statistical problems common to data used for ecological inferences, i.e. ecological data usually having high levels of heteroskedasticity, as well as other statistical normality issues.

A well-cited solution has shown in smaller geographic parameters such as neighborhood units, the inference problem is circumventable under a number of conditions, from which follows:

(1) The model should be scientifically validated extensive collections of real aggregate data.
(2) All components of the model are largely verifiable in aggregate data, thus while information is lost in aggregation some observable implications of model should remain in aggregate data.
(3) The assumptions on which the model concludes should remain robust to aggregation bias.
(4) The model should correct statistical problems that affect ecological inference (especially heteroskedasticity in aggregate data) and include extensive known information in the model.
(5) Stating the uncertainty of one's conclusions is especially important in ecological inference.
(6) The estimates should be accurate of the cross-tabulation cells at different geographic scales and should allow precinct-level parameters to vary in order to avoid geographic selection bias.
As such, the solution to the ecological inference problem is more or less a solution to the "modifiable areal unit problem", which occurs if widely varying estimates result when most methods are applied to alternate reaggregations of the same geographic units (King, 1997).

Solution 1, 2, and 3 should have been verified in this chapter and the previous chapter. Meanwhile, solution 4 is related to methodology and is discussed in Chapter 4, while solutions 5 and 6 are related to estimates and future research, and are discussed in the Chapters 5 and 6.

4 Methods

The main methods used is analysis of interactions in fixed effects OLS regression of panel data. Two models are used to analyze the hypothesized two-way and three-way interaction effects. This chapter covers the methodology of specifying and testing interaction effects in general and with fixed effects, followed by method, calculation, and interpretation of the moderation. All calculations are done in Stata. The final section covers model specification and selection.

4.1 The Approach

4.1.1 Interaction Effects

This section covers the concept of moderation, or interaction effects, and describes how moderator effects are tested and interpreted for the model types which are used in this analysis (Aiken & West, 1991; Dawson, 2014; Dawson & Richter, 2006), focusing on the methodology of two-way interactions, three-way interactions, and curvilinear interactions.

Testing for two-way interactions

Generally, an interaction is considered the measurement of how the effect of an independent variable changes with the size of another moderator variable. The simplest moderation form is a two-way interaction, where the relationship between an independent variable X and a dependent variable Y changes according to the value of a moderator variable Z. The statistical test for a two-way interaction involves an ordinary least squares (OLS) regression where the dependent variable Y is regressed on the interaction term XZ and the main effects X and Z. Including the main effects in the equation is essential, otherwise the results are uninterpretable. First, a simple test of a linear X-Y relationship is given by the regression equation of Y on X:

$$Y = \alpha + \beta_1 X + \varepsilon$$

where α is the intercept, β_1 is the coefficient of X, and ε is the residual. The two-way interaction expands to include the interaction term XZ created by multiplying X and Z together:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ + \varepsilon$$

The coefficients β_1 and β_2 determine whether there is any independent main effect of X or Z, but only the coefficients of the interaction term β_3 determines whether moderation is observed, and thus if Z is a statistically significant moderator of the linear relationship between X and Y. Normally, mean-centering (where the mean is subtracted from its variable) is advisable for continuous predictors (as is used in this analysis), since it eases interpretation of moderations, as the interpretation of the X coefficient is the relationship between X and Y when $Z = 0$.

Testing for three-way interactions

The moderating effects of a variable may itself depend on the values of another moderator, i.e. the relationship between net immigration and average income is assumed to be moderated by access to green spaces, while the moderating effects of green access may itself depend on the values of another moderator, such as income inequality, termed a three-way interaction. Statistically, it is represented by an extension to the two-way equation:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 WZ + \beta_7 XZW + \varepsilon$$

where W is a second moderator. The equation requires inclusion of the main effects of each of the three predictor variables, the three two-way interaction terms between each variable pair, as well as the three-way interaction term. All parts are required for meaningfully interpretation. The significance of the three-way interaction term β_7 coefficient determines if the moderating effect of the variable Z on the X–Y relationship is itself moderated by the other moderator W.

Testing for curvilinear effects in interactions

Sometimes non-linear effects are expected in moderation. Even if a curvilinear effects is not hypothesized it may still be useful to check if one exists, as linearity of the model is one of the assumptions for regression analysis. Insignificant effects in a linear model may miss an untested significant curvilinear effect where the X–Y relationship is significant at a particular value thresholds of X, which may be of interest to the analysis. Non-linear effects can be modeled in different ways, such as logarithmic, exponential, trigonometric, or reciprocal effects. This models the X-Y relationship by accounting for various effects such as curvilinear or U-shaped relationships, or where effects of X on Y vary more at higher, or lower, values of X. Statistically, a simple extension of the two-way regression is to include a quadratic element:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 Z + \beta_4 XZ + \beta_5 X^2 Z + \varepsilon$$

where X^2 is the independent variable X squared. Testing if Z moderates the relationship between X and Y is slightly more complicated. The significance of the coefficient β_5 shows if the curvilinear portion of the X–Y relationship is altered by the value of Z, and as such, if the form of the relationship is altered. Yet, it does not show if the strength of the relationship between X and Y is changed by Z. Testing the latter requires a joint test of coefficients β_4 and β_5 by using an F-test between the complete moderated quadratic regression model and an unmoderated quadratic model (excluding XZ and X^2Z), where H=0 is that the excluded variable coefficients equal zero. Testing yields a prob > F = 0.0318, and H=0 is rejected at 95% confidence value. Curvilinear three-way interactions are similarly calculated as a quadratic three-way extension:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 Z + \beta_4 W + \beta_5 XZ + \beta_6 X^2 Z + \beta_7 XW + \beta_8 X^2 W + \beta_9 ZW + \beta_{10} XZW + \beta_{11} X^2 ZW + \varepsilon$$

Here β_{11} determines if the curvilinear three-way interaction between the variables is significant. Curvilinear relationships between independent and dependent variables are not uncommon, which can normally be comprehensively captured by a quadratic regression. However, stronger correlations between X and Z impose an increasing risk that the true curvilinear X-Y relationship is instead erroneously picked up as an interaction between X and Z. Examining the analysis' variables suggests the presence of a curvilinear x–y relationship. Yet, testing yields a -0.0434 x-z correlation, which suggests a curvilinear interaction effect of a quadratic x-y relationship.

4.1.2 Fixed Effects vs. Random Effects

Cross-sectional heterogeneity in panel data is generally considered the normal state of affairs, as the influence of the unmeasured variables that determine *Y* yield different unit intercepts. The OLS may be biased unless the influence of these omitted variables is uncorrelated with the included explanatory variables. Usually causal estimation can be improved by modeling the presence of a different intercept for each cross-sectional unit via two different methodologies. This section discusses methods for exploiting the features of panel data to study causal effects, focusing on the choice between utilizing Fixed Effects (FE) and Random Effects (RE) models.

FE estimation is solely based on variation within units, thus automatically controlling for all observable and unobservable unit-specific characteristics (Allison, 2009; Wooldridge, 2010). The FE approach allow estimating causal effects in analyses of units measured over time, removing the effects of confounding time-invariant causes, measured and unmeasured, allowing the FE to alleviate omitted-variable bias even in a less-than-fully-specified model. FE models are especially useful where important causes of *Y* are hard to measure and tend to change slowly or immeasurably over time, a common situation in nonexperimental research (Firebaugh, Warner & Massoglia, 2013), i.e. the moderators green service area and gini index. FE is defined by an ordinary least squares (OLS) estimation on unit-mean centered data, which improves the potential of causal interpretations for the estimator (Gangl, 2010; Morgan & Winship, 2014). FE regressions are routinely utilized in empirical research, especially for panel data analysis (Young & Johnson, 2015). Many scholars have addressed the analytical properties of the FE estimator (Baltagi, 2005; Brüderl & Ludwig, 2014; Cameron, Trivedi & Trivedi, 2005), dealt with different approaches to its specification in regression frameworks (Andreß, Golsch & Schmidt, 2013; Firebaugh, Warner & Massoglia, 2013; Mundlak, 1978), theory-into-practice problems (Halaby, 2004; Plümper, Troeger & Manow, 2005), and inferential problems (Bell, Fairbrother & Jones, 2019). The methodology of FE is shown stepwise here (Firebaugh, Warner & Massoglia, 2013), Starting with standard regression model, which can be expressed in the following generic form, $\beta_X = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Y_i is the value of the dependent variable for the i^{th} unit, α is an intercept, β is a row vector of regression coefficients, *X* is a column vector of the *K* causes of *Y*, and ε is a random disturbance. Assuming that ε has a mean of zero and constant variance, and is uncorrelated with the variables in *X*, the equation is restated to distinguish the measured (*X*) and unmeasured (X^*) causes of *Y*:

$$Y_i = \alpha + \beta X_i + \beta^* X^*_i + \varepsilon_i$$

The vector *X* includes only a subset of the causal variables; the unmeasured causes are in vector X^*. Because some of the causal variables are missing, the regression of *Y* on *X* will generally result in biased estimates of the coefficients in β. However, with panel data the FE approach can alleviate the effects of confounding variables without measuring them, even without random assignment. This can be done by first amending the previous equation to represent an analysis based on panel data with $t = 1, 2, \ldots, T$ measurements for each unit:

$$Y_{it} = \alpha_t + \beta X_{it} + \beta^* X^*_i + \beta^{**} X^{**}_{it} + \varepsilon_{it}$$

By modeling different intercepts for each point in time, the term α_t allows for period effects that change the response variable by the same amount for each unit, permitting values to change for the i^{th} unit over time. β in the equation have no subscript t, as coefficients are time constant. Critically, the unmeasured causes are subdivided into time-variant (X_{it}) time-constant (X_i) Hence, for each of the stable unmeasured causes X_i, the product $\beta^* X^*_i$ is constant over time. Subsequently, the sum of the products is also time-constant for individuals, in other words, $\beta^*X^*_I = \beta^*_1X^*_{i1} + \beta^*_2X^*_{i2} + \cdots + \beta^*_iX^*_{iP} = \mu_i$, where μ_i is a constant for the i^{th} individual. Substituting μ_i for $\beta^*X^*_i$ yields the following equation:

$$Y_{it} = \alpha_t + \mu_i + \beta X_{it} + \beta^{**}X^{**}_{it} + \varepsilon_{it}$$

If all the unmeasured causes are time-invariant, then X^{**}_{it} is empty by definition, and the previous equation can be reduced to yield the foundational FE model:

$$Y_{it} = \alpha_t + \mu_i + \beta X_{it} + \varepsilon_{it}$$

As such, the FE model allows for both period-specific (α_t) and unit-specific (μ_i) fixed effects. The key term μ_i that varies across persons but is constant for each person over time then captures all relevant differences between individuals that time-invariant and unaccounted for by the other independent variables in the model. With panel data having multiple observations for each individual there are sufficient degrees of freedom to include dummy variables for μ_i. Hence, one common way to estimate a fixed effects model is with unit-specific dummies, which remove the stable effects (constant β) of time-invariant unmeasured causes (constant X_i).

As a result, the FE approach is less prone to bias because its assumptions about unmeasured causes are more realistic than the assumptions that is usually needed about those causes. Importantly, FE does not remove the biasing effects of time-varying confounders, so the key assumption that unmeasured causes are constant must still be met. However, the FE model has potential drawbacks: Including all individual dummies incurs a high loss in degrees of freedom. Additionally, since FE models remove the effects of all time-invariant causes the standard FE model is unable to estimate the effects of time-invariant measured causes. However, none of the included variables are time-invariant, thus this issue is inconsequential for this analysis. Finally, the focus on within-unit variance in FE also reduces its statistical power, since the μ_i term is an ignorance term, that is, it is a fitted value for each unit that reflects unit differences without indicating why the units are different. Although this makes FE effective in alleviating the confounding effects of unmeasured time-invariant causes, it involves loss of information.

Hence, the FE method is often less efficient than other estimation methods that are based on between-unit variance as well as on within-unit variance, such as the RE method. The RE method is the primary alternative method to estimating causal effects with panel data. The random effects approach treats the individual-specific effect as randomly varying, and unlike the FE method, RE makes use of between-unit as well as within-unit variance. In short, RE is computed using generalized least squares (GLS) to calculate a data transformation, creating a spherical variance-covariance matrix, then performing OLS on the transformed data. Potentially RE can yield more powerful hypothesis tests and narrower confidence intervals by retaining time-invariant variables, while saving degrees of freedom by not using unit dummies. Yet, it is more vulnerable to omitted-variable bias from unmeasured time-invariant causes.

When there is no positive correlation between X and intercept of individuals, OLS, FE, and RE estimators are all unbiased, but RE is most efficient. If a positive correlation between X and the intercepts of individuals is found, OLS and RE estimators are biased, but the FE is unbiased. Generally the RE estimator is only used whenever there is confidence that the model's composite error is uncorrelated with the explanatory variables (Kennedy, 2008). FE estimators are used in this analysis based on the Hausman test, which is covered later in this chapter.

4.1.3 Interactions in Fixed Effects models

The specification of interactions between continuous time-varying variables in Fixed Effects estimators follows the guidelines of previous work (Giesselmann & Schmidt-Catran, 2022; Kühhirt, 2012). In standard longitudinal regressions an interaction is specified by including the product of the original variables (Allison, 1977; Jaccard & Turrisi, 2003):

$$Y_{it} = \alpha_{\mathrm{t}} + \mu_i + \beta_X X_{it} + \beta_Z Z_{it} + \beta_{XZ}(X_{it}Z_{it}) + \varepsilon_{it}$$

Where β_{xz} measures the interaction between X and Z; β_x and β_z measure main effects at reference values of Z and X; and $\varepsilon_{\mathrm{it}}$ is the error term, which may include both unit-specific and time-specific components. The standard specification of an interaction in FE models is to treat the interaction term as a variable and demean it accordingly, with each realization of the product subtracted by its unit-specific mean before entering the regression:

$$Y_{it} - \bar{Y}_{i.} = \beta_X(X_{it} - \bar{X}_{i.}) + \beta_Z(Z_{it} - \bar{Z}_{i.}) + \beta_{XZ}\big(X_{it}Z_{it} - \overline{(XZ)}_{i.}\big) + \varepsilon_{it(fe)}$$

This specification for interactions in FE regression has been widely used in empirical practice (Killewald & Gough, 2013; Schofer & Longhofer, 2011), usually dubbed a within-estimation (Abendroth, Huffman & Treas, 2014; Kühhirt, 2012; Oesch & Lipps, 2013). Introduced as a desirable specification in methodological discourses, it is numerically equivalent to an OLS interaction estimation with unit dummy variables (Schunck, 2013), and is computed by default in statistical programs such as Stata (Cameron & Trivedi, 2010).

Yet, this strategy does not yield a within estimator of the interaction, only for the main effects. In the previous equation the FE estimator β_{XZ} exploits the between-unit differences in effects of Z, X, or both Z and X since the FE interaction estimator picks up unit-specific effect heterogeneity of both variables. This is contrary to the common view that FE estimates use only within-individual differences by dropping information about individuals differences (Allison, 2009; Halaby, 2004). As such, interactions estimated with standard FE may be subject to bias if both interacted variables vary within units, caused by the effects of heterogeneity across units if the effect of X on Z is influenced by two variables: observed time-varying Z_1 and unobserved time-constant Z_2. Using not just within-, but also between-unit differences in Z_1, the standard FE interaction estimator of β_{XZ} includes the moderating properties of any unobserved unit-specific characteristic correlated with z_1. Generally, the standard FE interaction estimator assumes that unit-specific specified and unobserved variables do not moderate the effect of X and uncorrelated with moderators Z and W. If this assumption holds, the standard FE estimator remains unbiased and can be used for interpretating the interaction. Otherwise, using an alternative double-demeaned within-unit interaction estimator can avoid the effect of unobserved heterogeneity (Giesselmann & Schmidt-Catran, 2022; Shaver, 2019). Considering which estimator is preferred in models with two time-varying interacted variables concerns the debate on trade-offs between consistency and efficiency of within-estimators. Some studies advocate cutting between-variation and using within-estimators in causal analysis (Brüderl & Ludwig, 2014; Halaby, 2004), while others justify the use of between-unit variation in the absence of time-constant unobservable confounders (Allison, 2009; Wooldridge, 2010). The issues above mainly concerns longitudinal microdata with cross-sectional heterogeneity, and should be less of a problem in aggregated panel data, which averages away heterogeneity. Given the ecological nature of the data in this analysis, the tested absence of omitted variables, as well as residual normality, the effects of heterogeneity on variables is assumed to be zero. Thus, it is assumed that no unit-specific specified and unobserved variables moderate the effect of net immigration (*x*) or correlate with moderators green service area *(z)* and gini index *(w)*. Hence, the standard FE estimator with included between-variation of the interactions should be unbiased and more efficient, and is therefore used for interpretation of interaction effects.

4.1.4 Interpreting Interaction Effects

Interpretation of interaction effects will take inspiration from the methodology "simple slopes" (Dawson, 2014; Dawson & Richter, 2006). From a significant interaction term the slopes of the lines are known to be significantly different from each other (Aiken & West, 1991), and the direction of the relationship is known. A significant interaction gives only the result that the association between *X* and *Y* differs according to the level of *Z* (and *W*), yet it is not entirely clear how it differs, e.g. a positive interaction coefficient suggests that it becomes more positive at higher values of *Z*, yet the size and nature of the effect is not clear, and is further difficult to probe if some coefficients are negative, or the standard deviations of *X* and *Z* are very different.

One way to overcome this is to plot the interaction effects visually for eased interpretation by calculating simple slopes of the moderation at different low and high levels of the predictor and the moderators, and then examine the specific *X*-*Y* relationship at these particular levels, and if the moderation is significant, which may be done via simple slope tests or other tests. Also, for a three-way interaction testing whether the difference between a pair of slopes is significant may be useful, as there are several possible pairs of lines for different combinations of the two moderators, and the interaction term alone cannot tell if each line pair is different. Additionally, testing whether a curvilinear *X*–*Y* relationship significant at a particular *Z* values requires distinguishing between testing if the curvilinear relationship a significant at that value and testing if there is any relationship at all at that value. As there is no hypothesis on the curvilinear nature of the *x*-*y* relationship at specific *z* values, testing the significance of curvilinearity is meaningless, hence only the interaction significance at each *z* value is needed.

For two-way models, slopes are calculated for predicted *X*-*Y* relation at different conditions of high and low values of *X* and Z, while three-way models calculate the predicted *X*-*Y* relation at different conditions of *Z* and *W*. The question is what values are meaningful to be plotted. A common method is using values at **one standard deviation above and below the mean.** However, this method may result in simple slope tests using arbitrary values of the moderator, and can be problematic as significance of the slope may depend on specific moderator values.

For the standard FE regression model, the most popular method for probing interactions is the **"pick-a-point" approach** (Rogosa, 1980), which plots and tests the conditional effect of the focal predictor at designated levels of the moderating variable (e.g., high, medium, and low), where these conditional effect estimates are commonly referred to as "simple slopes" (Aiken & West, 1991; Jaccard & Turrisi, 2003). However, choosing moderator values to plot the slopes demands theoretical justification (Bauer & Curran, 2005; Hayes, 2022), to which the literature review provided little specific insights. While higher values of *z* green service area are expected to increase the slope of the *x*-*y* relationship, the exact values where this moderation may affect the relation are not theorized. Lacking theoretical justification there are other options available for choosing which values of moderation for the simple slopes are to be tested.

A popular alternative to circumvents such arbitrariness is evaluating **regions of significance** (Aiken & West, 1991), which seeks to identify a region of values of Z where the *X*–*Y* relationship would be statistically significant. This may help understanding the *x*-*y* relationship of this sample, as it can indicate the values of green service area where immigration is more likely to be important for neighborhood wages. It should still be interpreted with caution, as there is nothing special about the identified region, which is merely the value in the particular dataset where the relationship is significant. Importantly, the size of the region of significance is dependent on the sample size, and the boundaries of the region do not represent estimates of any meaningful population parameters. Yet, if correctly interpreted, region of significance is more useful than the simple slope tests. Thus, regions of significance are used as a starting point for plotting the two-way interaction.

One approach is to testing the region of significance is the **Johnson–Neyman (J–N) technique** (Johnson & Neyman, 1936), which describes the variability about the estimates by obtaining regions of significance by calculating a critical value through the construction of confidence bands around the simple slopes of the moderation (Bauer & Curran, 2005; Preacher, Rucker & Hayes, 2007). Simple slopes and J-N techniques both rely on traditional null hypothesis testing logic. However, confidence intervals are generally considered to provide more information than hypothesis tests and it is increasingly recommended to use of confidence intervals in addition to or in place of hypothesis testing. Thus, regions of significance are commonly calculated using delta method standard errors, where the formula for a 100 × (1 − α)% CI for a simple slope is as follows:

$$CI_{\widehat{\omega}_1} \pm t_{crit} SE_{\widehat{\omega}_1}$$

As the formula for $SE_{\widehat{\omega}_1}$ relies on a particular *z* value, $CI_{\widehat{\omega}_1}$ varies as a function of the moderator. When $CI_{\widehat{\omega}_1}$ is plotted across all relevant values of *z*, the result is a pair of confidence bands (Preacher, Curran & Bauer, 2006). While extensions of the J-N technique allow testing continuous-by-continuous interactions (Finsaas & Goldstein, 2021) it is not easy to directly implement in Stata. Hence, a similar approach is utilized to identify the region of significance where moderator values are significantly different from zero, via continuously plotted confidence intervals around slopes of the interaction effect at all moderator values.

The approach is thus: For the two-way model to identify the region of significance for the moderation effect by calculating the slope confidence intervals of the *x* net immigration effect on *y* log wage at combinations of *x* and the moderator *z* green service area. Three-way models, due to the complex nature of multiple moderators, are advised to choose conditional values of one moderator and obtain regions of significance for the other moderator at those conditions

(Preacher, Rucker & Hayes, 2007). The moderation effect of interest is (z) green service area at different conditions of (w) Gini index, thus the moderated region of significance of the x-y relationship is calculated for all combinations of x and z, at conditional values of w.

Margins

To this end, analysis of margins is used as an approximation to calculate the regions of significance for the moderation effects, which can subsequently be visualized and interpreted (UCLA, 2022a, 2022b). *Margins* is a postestimation command in Stata to estimate and report margins, which is a statistic obtained from predictions of a fitted model calculated over a dataset in which some or all the covariates are fixed at values different than their actual value, and averaging or integrating over the remaining covariates. The margins command produces a variety of estimates, of which those important for this analysis are the margins of responses (a.k.a. predictive margins) and the margins of derivatives of responses (a.k.a. marginal effects). In short, marginal effects provide summaries of the models, while predictive margins provide unit-specific and sample average predictions from the models (Stata, 2021; Williams, 2012). Marginal effects can summarize the moderation effect and yield the region of significance, while the predictive margins can plot the significant predicted values of the moderation effect, thus both marginal effects and predictive margins are used for analysis to allow interpretation. First, marginal effects are calculated for the response surface of the x-y relationship at representative fixed moderator values to identify the region of significance for the effect. Hereafter, predictive margins are used to plot and visualize the slopes of the predicted changes in the x-y relationship at the representative moderator values within the region of significance.

Predictive margins are estimates of the response mean used when fixing some, but not all, predictors in the model at specified values, as is case here. As such, predictive margins are model summaries in the form of adjusted predictions, which allow extracting the equivalent of the main effects and interaction effects from a model. Thus, predictive margins provide predicted measures of changes in the response for change in a covariate, in other words, predictive margins predict the y value at the fixed values of the predictor and the moderators.

Marginal effects, on the other hand, are partial derivatives of the regression equation with respect to each variable in the model for each unit in the data, and refers to the derivative of y and the derivative of x. Derivatives are of interest because they are an informative way of summarizing fitted results, making the change in a response for a change in the covariate easy to understand and to explain. Marginal effects are referred to as average marginal effects when some covariates are not fixed, as is the case in this analysis, where only x, z, w, are fixed at different values. The average of the marginal effects over the observations thus measure the effect of the continuous predictor on the response mean. This obtains average marginal effects, which are simply the mean of these unit-specific partial derivatives over some sample. Specifically, average marginal effects of a continuous predictor at a representative observation estimates the slope of the mean response curve at that observation's setting of the predictors, and is computed as the partial derivative of the mean with respect to the predictor. Therefore, it can be interpreted as the instantaneous rate of change of the response mean at that point.

Marginal effects with the **dydx(*x*) at(*z*) at(**w**)** options are used for evaluating the marginal effects over the response surface to find the significance of slopes of the x-y relation at all modifier values, which allows identifying the region of significance for the moderated effects.

Using the **dydx()** option estimates marginal effects for specified covariates values of interest, by means of using **dy=d(varname)** as response variable. The formulas is: **dydx() = dy=dx** which can be interpreted as a change in *y* for a change in *x*. Importantly, **dydx()** is a rate, and all such interpretations are valid only for small (infinitesimal) changes in *x*, therefore the casual interpretation of **dydx()** is that it represents the response to a unit change. In other words, **dydx()** = *value* means that *y* increases with *x* at a rate such that, if the rate were constant, *y* would increase by *value* if *x* increased by 1. Where marginal effects **dydx()** calculates effects, predictive margins yields predictions, yet, calculating marginal effects also uses predictions by changing values by a small amount.

Applying the frequently used option **at()** with marginal effects makes it easier to understand how the response varies by exploring the nature of the response surface. The **at()** option calculates marginal effects at potentially representative specified values by replacing observed values with specified replacement values before calculating marginal effects. Margins controls the values in each **z** vector via the **marginlist** and the **at()** option, where the **at()** option sets model covariates to fixed values, temporarily setting *x* to that value for each observation in the dataset before computing any predictions. Calculations are made at the observational level and are then averaged. For predictive margins, the **at()** option can similarly specify the simple slopes at the representative values within the region of significance of *x* and *z* in the two-way model, and of *z* and *w* in the three-way model.

Additionally, the option **pwcompare** with marginal effects performs pairwise comparisons across the levels of factor variables from the model, and compares estimated marginal slopes and reports the comparisons as differences of margins along with significance tests or confidence intervals for the contrasts. This option can thus be utilized for to test for pairwise statistical significance of the slopes in the three-way interaction model. Significance test for pairwise differences may use the Bonferroni correction, which usually used for discovered rather than an expected interactions and is more conservative. Since the interaction is hypothesized and the interaction term is weakly significant, the Bonferroni correction can accommodate this uncertainty in the test.

Finally, the standard errors and confidence intervals produced by margins are based on the delta method applied to the VCE of the estimates, and treat the covariates at which the response is evaluated as given or fixed, which is appropriate when the **at()** option is used to fix the covariates. Appendix D elaborates on statistical formulas for the computations.

4.2 The Model

Before regression, both the existence and the form of an interaction effect must be predicted, specifically if a moderator increases or decreases the association between two other variables, as priori hypotheses is a requirement for meaningful significance testing of moderation (Dawson, 2014). As such, this section overs the expected signs and effects of the moderations in both the two-way and the three-way model, and follows up with the methodology applied to specification and selection of parsimonious models based on predictions and goodness of fit.

4.2.1 Two-Way Model

For two-way interactions it is sufficient to simply state the direction of the interaction, although hypothesizing whether the main *x-y* effect is positive, negative, or null at high and low values of the moderator is necessary for generating meaningful simple slope tests at these values. Based on theory the hypothesized FE two-way interaction model of the moderating effect of green service on the relationship between immigration and log wage is:

$$y_{it}\,\text{log wage} = \alpha_t + \mu_i + \beta_1 x_{it}\text{net immigration} + \beta_2 z_{it}\text{green service area} + \beta_3(x_{it}z_{it}) + \beta_k c_{it} + \varepsilon_{it}$$

where $\beta_k c$ are the added control variables. The coefficient of the interaction term β_3 is expected to show positive signs, and as such the main effect of (x) net immigration should have a positive impact on the linear predictions of on (y) log wage at higher values of (z) green service area, as access to green space may attract wealthy residents to the neighborhood and increase wages. The coefficient for the *x-y* relationship term $\beta_1$is similarly expected to be positive, whereas the *z-y* relationship term β_2 is expected to be positive, but statistically insignificant, as per theory. However, the linear interaction may be statistically insignificant since the expected relationship between net immigration and log income is suspected of being curvilinear, as seen in Figure 4. To test for non-linear moderation uses the following FE quadratic two-way interaction model:

$$y_{it}\,\text{log wage} = \alpha_t + \mu_i + \beta_1 x_{it}\text{net immigration} + \beta_2 x_{it}^2 + \beta_3 z_{it}\text{green service area} + \beta_4(x_{it}z_{it}) + \beta_5(x_{it}^2 z_{it}) + \beta_k c_{it} + \varepsilon_{it}$$

where the quadratic interaction β_5 is expected to show statistically significant positive signs. The expected main effect of the moderation is similar to the linear interaction, however considering the nonlinearity of *x* covering negative and positive values, the main moderation effect is expected to be positive at both increasing negative and increasing positive values of *x*.

4.2.2 Three-Way Model

For three-way interactions priori hypothesis is more complex, and requires hypothesizing how the slopes of the interaction should differ for different value combinations of the moderators. The hypothesized FE two-way interaction model of the moderating effect of green service area and gini index on the relationship between net immigration and log wage is thus as follows:

$$y_{it}\,\text{log wage} = \alpha_t + \mu_i + \beta_1 x_{it}\text{net immigration} + \beta_2 z_{it}\text{green service area} + \beta_3 w_{it}\text{gini index} + \beta_4(x_{it}z_{it}) + \beta_5(x_{it}w_{it}) + \beta_6 z_{it}w_{it} + \beta_7(x_{it}z_{it}w_{it}) + \beta_k c_{it} + \varepsilon_{it}$$

where the interaction term β_7 is expected to be statistically significant. Theory suggests that the moderating effect of green service area at higher levels of gini index income inequality should exacerbate the effect of socioeconomic displacement, where net negative migration of poor residents should increase the average wages of a neighborhood. Conversely, the higher levels of green service area at lower income inequality levels should have a more positive moderating effect on the relation between net immigration and average wages, as there may be fewer poorer migrants to displace, and thus a stronger effects of inward migration of the richer. Testing for a quadratic three-way interaction with a curvilinear predictor *x* uses the following:

$$y_{it}\ \text{log wage} = \alpha_t + \mu_i + \beta_1 x_{it}\text{net immigration} + \beta_2 x_{it}^2 + \beta_3 z_{it}\text{green service area} + \beta_4 w_{it}\text{gini index} + \beta_5(x_{it}z_{it}) + \beta_6\left(x_{it}^2 z_{it}\right) + \beta_7(x_{it}w_{it}) + \beta_8\left(x_{it}^2 w_{it}\right) + \beta_9(z_{it}w_{it}) + \beta_{10}(x_{it}z_{it}w_{it)} + \beta_{11}(x_{it}^2 z_{it}w_{it)} + \beta_k c_{it} + \varepsilon_{it}$$

The interaction term β_{11} is not expected to be significant, since the opposing effects of the two slopes at different moderator conditions would likely tend toward 0 impact on the x-y relation in a nonlinear interaction, due to the previously discussed nonlinear nature of x covering positive and negative values. As such, the term may show either positive or negative signs, but in case of non-significance further interpretation may not be meaningful.

4.2.3 Model Specification

Model Selection

In order to select a parsimonious model, a combination of two main approaches are conducted, approximating Average Economic Regression (AER) and The Test, Test, Test (TTT) methods (Kennedy, 2008). The AER approach is used as the initial method for model selection, beginning with a hypothesized specification of a simple model derived from theory and thereafter proceeding with testing-based forward model selection to identify a general model. Considering the fact that the hypothesized specification is only partially derived from theory, while investigating the nature of a known phenomenon at specific parameters, the approach of this analysis thus differs from the standard AER approach. Statistical diagnostic tests are interpreted not only in terms of estimation problems, but also for potential misspecification. Hereafter the TTT approach is similarly approximated in order to discover tenable models. First, the initial specification is made more general than the expected model specification, whereafter testing of various restrictions is undertaken to simplify the general specification, i.e. backwards testing-based selection, which avoids selection bias from pure forward selection. Models are continually respecified until diagnostic tests can conclude that the model is congruent with the evidence, for which there are five main criteria, where models should be:
(1) Data-admissible, thus not capable of producing predictions that are not logically possible;
(2) Theory-consistent, following expectations of the economic theory from which it is derived;
(3) Parameter-consistent, able to adequately predict observations not used in the specification;
(4) Data-coherent, with random distribution of residuals to ensure regularity is not excluded;
(5) Encompassing, where rival models have no information able to improve the current model.
Throughout both processes, the models are subjected to a variety of diagnostic tests, which in unison are considered to identify the most parsimonious models. Diagnostics tests include:

- Evaluating coefficients for correct signs, expected significant *t* values, and consistency.
- Multicollinearity is tested for before applying FE, as it is not testable with the FE estimator.
- Heteroskedasticity is tested graphically, and autocorrelation is tested statistically.
- Normality of residuals is tested statistically and graphically.
- R^2 values are considered only casually for model selection, as additional variables increase the statistic, and the adjusted R^2 is not applicable when using robust standard errors.
- The Ramsey Regression Equation Specification Error Test (RESET) is utilized to check for misspecification of functional form of variables.
- The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used as algorithmic selection to guide the selection of models, in unison with other diagnostics.

Estimator and Standard Errors

In order to test whether FE or RE is the correct specification, a variant of the Hausman test (Hausman, 1978) is generally utilized to comparing the results from the fixed and random effects models to determine if there is sufficient evidence to reject the null hypothesis that the unobserved individual differences are orthogonal to the regressors in the model. If the null hypothesis cannot be rejected, then it is assumed that the unobserved heterogeneity is uncorrelated with the regressors, and thus that the RE estimates are consistent. Alternatively, if the null hypothesis is rejected, the RE estimates is inconsistent, and must be rejected in favor of the FE estimates (Firebaugh, Warner & Massoglia, 2013). Hence, the Hausman test is performed on both interaction models to select between FE or RE (Table 5), where in both cases the null hypothesis rejected at a 1% confidence level, and FE estimator is used.

Table 5. Hausman test for H=0: Difference in coefficients not systematic

Three-way interaction model	Two-way interaction model
chi2(1) = 43.56, Prob > chi2 = 0.0000	chi2(1) = 14.11, Prob > chi2 = 0.0002

Heteroscedasticity and autocorrelation are common issues in panel data (Saeed et al., 2018), where heteroskedasticity can make ordinary OLS methods inefficient via biased error variance, and autocorrelation can lead to inconsistent estimates in dynamic panels. In use of panel data, clustering should be considered when heteroskedasticity and/or autocorrelation is suspected. Cluster robust standard errors account for heteroskedasticity in the unexplained variation, i.e. if the amount of variation in the outcome variable is correlated with the explanatory variables, robust standard errors account for this correlation, which thus also accounts for autocorrelation. Since the panel data consists of a comparatively large N=67 and a very small T=3 with time gaps, few tests are significantly able the Portmanteau IS-test can be used for panel data with gaps and significant N>T to check for autocorrelation in the FE, by testing if any off-diagonal element of the autocovariance matrix is a vector of error terms (Born & Breitung, 2016; Feng et al., 2020; Inoue & Solon, 2006). The test does not work in the presence of heteroscedasticity. Importantly, the significantly small T=3 of the panel data does not allow any statistical test to significantly check for groupwise heteroskedasticity. However, graphical checks by plotting the combined residuals against the fitted values indicates inconstant variance of residuals (Figure 7). Hence, the Portmanteau IS-test and similar cannot be used. Due to the presence of heteroskedasticity and possible autocorrelation cluster-robust standard errors are utilized.

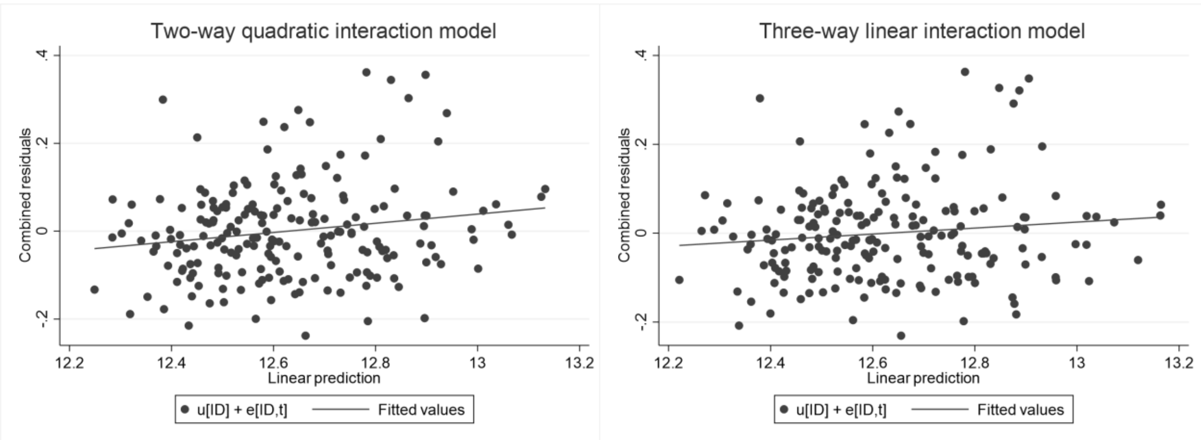


Figure 7. Graphical check for heteroskedasticity – combined residuals against linear prediction

5 Empirical Analysis

This chapter covers the analysis of the two-way and three-way interactions of the relation between net immigration and wage at moderation of green service area and income inequality at the neighborhood level. The analysis first examines the regression results and then interprets the interactions via the marginal effects and predictive margins based on moderator values.

5.1 Results

5.1.1 Two-Way Interaction Results

The quadratic two-way FE regressions produce statistically significant interaction terms throughout the specification process and is stable with additional control variables (Table 6). In the linear two-way model the interaction term is consistently insignificant (Appendix E).

Table 6. Comparison of regression results - quadratic two-way models – fixed effects

	Restricted	*Simple*	***Full (selected)***
Net immigration	9.676*	5.980	2.584
Net immigration^2	-171.318*	-96.491*	-61.954**
Green service area	0.377	0.054	0.028
Net immigration # Green service area	-22.031*	-12.348*	-6.497*
Net immigration^2 # Green service area	**326.181***	**172.220***	**101.210****
Gini index	.0385**	0.014***	0.015***
N 0-19- years %		727*	0.837**
N 79-99+ years %		1.316***	1.596***
Higher education %		1.385***	1.792***
Lower education %		-0.633**	-0.849***
Unemployed %			-1.000***
Women %			-2.162***
Non-western ethnicity %			1.477***
Constant	11.281**	11.855***	11.641***
R^2 within	0.6738	0.907	0.9428
R^2 between	0.5287	0.747	0.8366
R^2 overall	0.5619	0.776	0.8540
Correlation (u_i, Xb)	-0.3612	0.1799	0.1636
AIC, BIC	-593, -570	-836, -803	-863, -821

*Standard errors: *** p<0.01, ** p<0.05, * p<0.1*

The quadratic interaction is generally stable throughout the model specifications in models which pass diagnostics tests. The parsimonious model should be congruent with the evidence, achieving good explanatory power under limited, yet theoretically required predictor variables. Hence, the quadratic interaction margins may be casually interpreted for the first part of the research question, which lay the foundation to contrast with analysis of the three-way model. Based on the beforementioned methodology the *margins* command is used in Stata to compute marginal effects for response variable *y* log wage and predictor *x* net immigration while holding the moderator *z* green service area constant for virtually all values of *z* at increments of 0.001 between z = 0.05 to 0.85, which thus covers the upper and lower boundaries of the moderator.

. margins, dydx (*net immigration*) at (*green service area* = (0.05 (0.001) 0.85))

This obtains slope coefficients and confidence intervals of the marginal effects for each slope (Appendix F, generalized to increments of 0.1). The statistical output indicates that the moderating effects of green service area are only significant at values 0.44 and higher. Thus, the moderation slopes are then recalculated for the identified region of significance (Figure 8).

. margins, dydx (*net immigration*) at (*green service area* = (0.044 (0.001) 0.85))

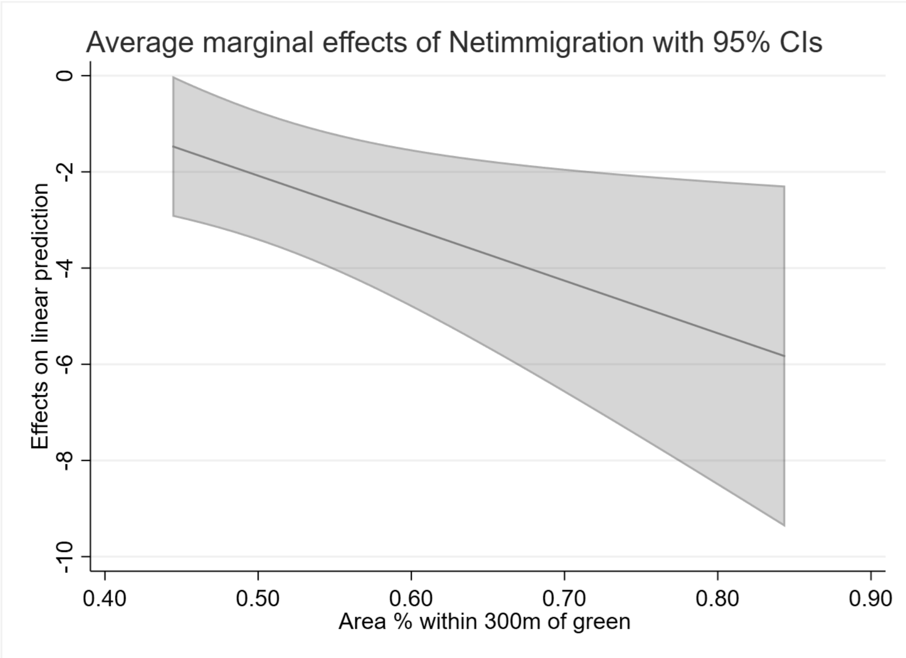


Figure 8. Average marginal effects of net immigration on wages at values in the region of significance for green service area.

The outputs show the average marginal effects of net immigration on the average wages levels being negative and decreasing with higher levels of green service area. From this it seems that access to green spaces within 300 meters has a negative effect on attraction of richer residents, which is the opposite effect than what was hypothesized based on theory. The marginal effects, however, show only the moderation on the average effect of net immigration. Considering the nonlinear nature of the interaction, it may be informative to plot the predictive margins as both linear and nonlinear to examine the differences in the moderation, calculated as (Figure 9):

. margins, at (*net immigration* = (-0.05   0.1) *green service area* = (0.44 (0.1) 0.85))
. margins, at (*net immigration* = (-0.05 (0.025) 0.1) *green service area* = (0.44 (0.1) 0.85))

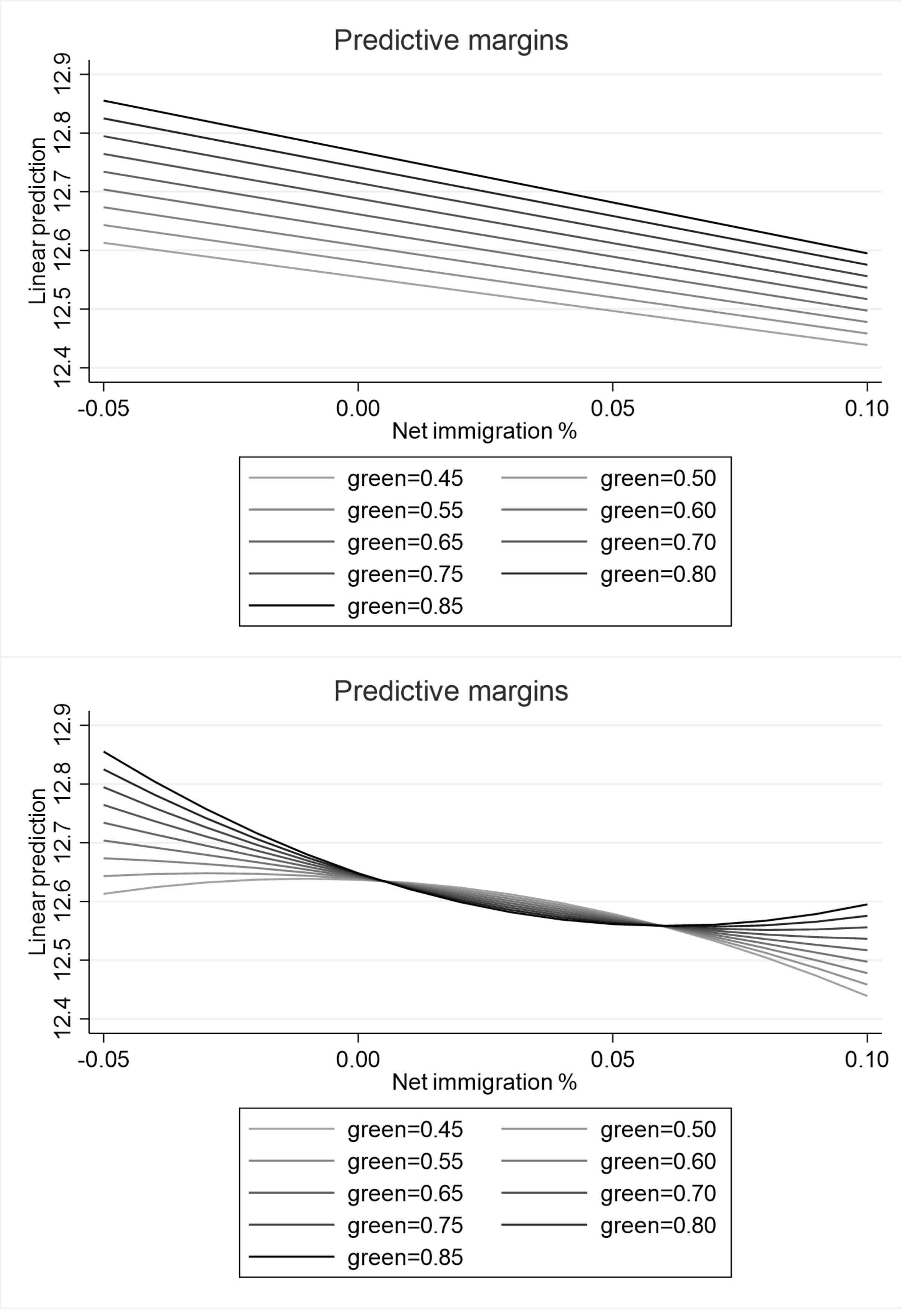


Figure 9. Two-way interaction – simple slopes of predicted margins in the region of significance - linear(upper) and nonlinear(lower).

While simple slopes may have some merit in the moderation effect, it may be more illustrate to visualize the nonlinear interaction using *contours* (Huber, 2022; Rising, 2011) (Figure 10):

. twoway contour y (log wage) (x) net immigration z (green service area)

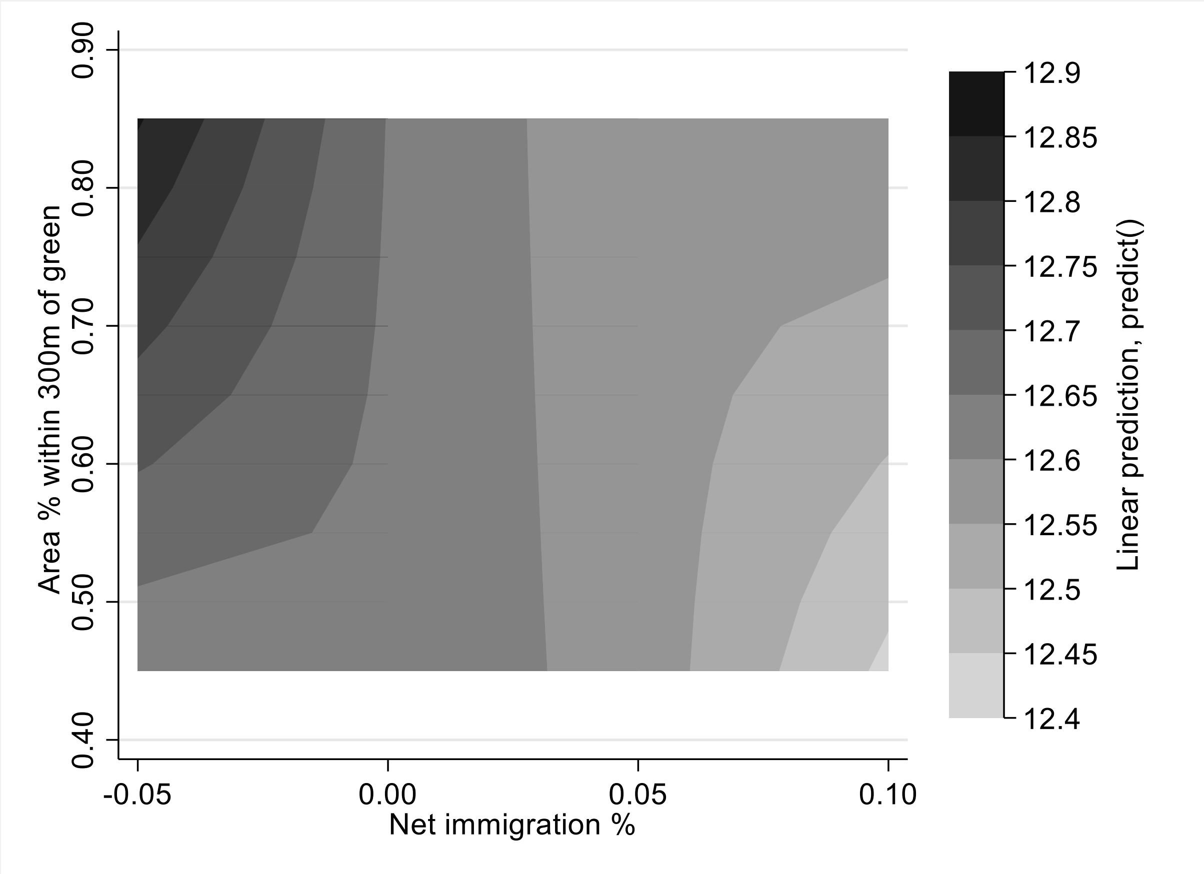


Figure 10. Two-way quadratic interaction - contour of predicted margins (region of significance).

The figures seem to portray a somewhat more nuanced image of the moderating effect of access to green service area. The unmoderated relation predictor net immigration has a curvilinear negative relation with log wage (Appendix H). Yet, the moderation effect of green service area generally has a positive impact on the average relationship between immigration and log wage. The linear predictive margins indicates that on average higher levels of green service area increase the positive impact on the relation, i.e. decreases the strength of the negative relation. When measured on the nonlinear relation between net immigration and log wage, the positive moderation effect of higher levels of green service area is strongest mainly at net negative values and to a lesser extent at higher positive values of net immigration. The average marginal effects of the moderation are interpreted as the average rate of change of the response for one (infinitesimal) unit change in the predictor for each moderation level. Examining the average marginal effects of the moderating effect (Table 7), the average rate of change in the response of log wage to one unit change in immigration decreases by 1.55 over increasing green moderation levels (between individual neighborhoods over time). Of note, the negative values of net immigration are positively related with the response on log wage, and thus the moderation should be interpreted as positive when moving away from zero. Predictive margins (Table 8) show this by deriving a ca. moderated 1.92% higher predicted response value at high negative immigration, and ca. 1.26% higher response at high positive immigration levels.

Table 7. Average marginal effects: net immigration on log wage at green service area moderation – region of significance

Net immigration at:	dy/dx	Delta-method std. err.	z	P>z	[95% conf.	interval]
green = 0.45	-0.860	0.436	-1.970	0.049	-1.715	-0.005
green = 0.55	-1.096	0.421	-2.600	0.009	-1.921	-0.271
green = 0.65	-1.532	0.506	-3.030	0.002	-2.523	-0.541
green = 0.75	-1.969	0.684	-2.880	0.004	-3.309	-0.629
green = 0.85	-2.406	0.902	-2.670	0.008	-4.173	-0.638

Table 8. Predictive margins of net immigration on log wage at green service area moderation – region of significance

Linear prediction	*_at*	*Margin*	*Delta std. err.*		*P>z*	*[95%*	*interval]*
1: net immigration=-.05	*green=.45*	12.61292	.0650559	193.88	0	12.48541	12.74042
2: net immigration=-.05	*green=.55*	12.67350	.0585397	216.49	0	12.55876	12.78823
3: net immigration=-.05	*green=.65*	12.73408	.0677684	187.91	0	12.60125	12.8669
4: net immigration=-.05	*green=.75*	12.79466	.0879161	145.53	0	12.62235	12.96697
5: net immigration=-.05	*green=.85*	12.85524	.1133002	113.46	0	12.63318	13.0773
6: net immigration=0	*green=.45*	12.63695	.0071423	1769.32	0	12.62295	12.65095
7: net immigration=0	*green=.55*	12.63974	.0047881	2639.82	0	12.63036	12.64913
8: net immigration=0	*green=.65*	12.64254	.0121487	1040.65	0	12.61873	12.66635
9: net immigration=0	*green=.75*	12.64533	.0206156	613.39	0	12.60493	12.68574
10: net immigration=0	*green=.85*	12.64813	.0292432	432.51	0	12.59081	12.70545
11: net immigration=.05	*green=.45*	12.57893	.0165144	761.70	0	12.54656	12.6113
12: net immigration=.05	*green=.55*	12.57455	.0142136	884.69	0	12.54669	12.60241
13: net immigration=.05	*green=.65*	12.57016	.017043	737.55	0	12.53676	12.60357
14: net immigration=.05	*green=.75*	12.56578	.0231958	541.73	0	12.52032	12.61124
15: net immigration=.05	*green=.85*	12.56139	.0307369	408.67	0	12.50115	12.62164
16: net immigration=.1	*green=.45*	12.43887	.0778524	159.78	0	12.28628	12.59146
17: net immigration=.1	*green=.55*	12.47791	.0710062	175.73	0	12.33874	12.61708
18: net immigration=.1	*green=.65*	12.51695	.0680253	184.00	0	12.38362	12.65028
19: net immigration=.1	*green=.75*	12.55599	.0694094	180.90	0	12.41995	12.69203
20: net immigration=.1	*green=.85*	12.59503	.0749172	168.12	0	12.4482	12.74187

Hence, casual inference is that higher levels of access to green within 300 meters is associated with increased neighborhood wages through migration, especially at negative migration levels, which follows the hypothesized expectations of hedonic pricing and possible displacement of socioeconomically weaker residents. Thus, analysis of the three-way moderation is warranted.

5.1.2 Three-Way Interaction Results

The FE regression yield a weak statistically significant linear three-way interaction (Table 9) for the predictor net immigration and the two moderators green service area and Gini index. The parsimonious model should be congruent with evidence, achieving good explanatory power under a theoretically required predictor variables while passing all diagnostics tests, yet the interaction term remains only weakly significant at a 90% confidence level in good models. Thus, the wider confidence interval warrants a slighter more casual interpretation of the effect.

Table 9. Comparison of regression results - linear three-way models – fixed effects

	Restricted	***Simple (selected)***	*Full*
Net immigration	-23.783*	-10.748	-7.693
Gini index	-.357	0.008	0.013*
Green service area	0.027**	-0.391	-0.884**
Net immigration # Gini index	42.009**	0.363	0.330
Net immigration # Green service area	0.722*	23.124	32.84001**
Green service area # Gini index	0.021	0.013	0.037***
Net immigration # Green service area # Gini index	**-1.394****	**-0.803***	**-0.967***
N 20-69 years %		-0.936***	-0.999***
Medium education %		1.267***	1.028***
Lower education %		-0.803***	-1.769***
Women %			0.152
Unemployed %			-1.174***
Danish ethnicity %			0.425**
Constant	11.682***	13.056***	12.608 ***
R^2 within	0.668	0.904	0.863
R^2 between	0.506	0.785	0.805
R^2 overall	0.546	0.805	0.818
Correlation (u_i Xb)	-0.373	0.2207	-0.495
AIC, BIC	-588, -561	-829, -796	-754, -711

*Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1*

Also, no significant quadratic three-way interaction was found, with the term being consistently insignificant in the specification process of models which pass diagnostics tests (Appendix H). In order to calculate the simple slopes the *margins* command is used to compute the average marginal effects of net immigration on the linear prediction over the values of green service at high and low values of gini index (one standard deviation), which should yield the moderating effect of gini index on the previously established two-way moderation:

. margins, dydx(*net immigration*) at(*green service area*=(0.05(0.1)0.85) *low gini, high gini)*

The output in Table 10 and Figure 12 shows the area of significance for the green service area as a moderator for at high and low values of Gini index: At high Gini index the values of green service area are significant from ca. 0.05-0.30 and from 0.60-0.85. At low values of Gini index, the relationship is unchanged from the two-way model, with green service area significant at values 0.45 and above. With the identified region of significance for the three-way moderation the predictive margins are calculated, as seen in Table 12 and in Figure 11:

. margins, at (*net immigration* = (-0.05(0.05)0.1) *green service area* = (0.05(0.2)0.85) *gini index* = (*low gini index, high gini index*))

Table 10. Average marginal effects of net immigration moderated by values of green service area at high and low gini index (1 std.dev.)

Net immigration at: High Gini index and		*dy/dx*	*Delta-method std. err.*	*z*	*P>z*	*[95% conf.*	*interval]*
1._at: green =	*.05*	3.864	1.462	2.640	0.008	0.998	6.729
2._at: green =	*.15*	3.013	1.245	2.420	0.016	0.573	5.453
3._at: green =	*.25*	2.162	1.033	2.090	0.036	0.137	4.188
4._at: green =	*.35*	1.311	0.832	1.580	0.115	-0.319	2.942
5._at: green =	*.45*	0.461	0.649	0.710	0.478	-0.812	1.733
6._at: green =	*.55*	-0.390	0.506	-0.770	0.441	-1.383	0.602
7._at: green =	*.65*	-1.241	0.444	-2.790	0.005	-2.111	-0.371
8._at: green =	*.75*	-2.092	0.494	-4.240	0.000	-3.059	-1.124
9._at: green =	*.85*	-2.943	0.629	-4.680	0.000	-4.176	-1.709
Low Gini index and							
1._at: green =	*.05*	0.526	2.734	0.190	0.847	-4.832	5.885
2._at: green =	*.15*	0.821	2.258	0.360	0.716	-3.605	5.248
3._at: green =	*.25*	1.116	1.793	0.620	0.534	-2.399	4.630
4._at: green =	*.35*	1.410	1.348	1.050	0.295	-1.232	4.052
5._at: green =	*.45*	1.705	0.952	1.790	0.073	-0.161	3.570
6._at: green =	*.55*	1.999	0.695	2.870	0.004	0.636	3.362
7._at: green =	*.65*	2.294	0.741	3.100	0.002	0.842	3.746
8._at: green =	*.75*	2.588	1.050	2.470	0.014	0.531	4.646
9._at: green =	*.85*	2.883	1.464	1.970	0.049	0.014	5.752

Table 11. Pairwise comparison of slopes in three-way model.

Net immigration at:	*Contrast dy/dx*	*Delta-method std. err.*		*Bonferroni*		*Bonferroni*
			z	P>|z|	[95% conf. interval]	
2 vs 1	-.3377191	1.546737	-0.22	1.000	-4.41841	3.742971
3 vs 1	.9406268	1.577193	0.60	1.000	-3.220413	5.101667
4 vs 1	-3.054861	1.384636	-2.21	0.164	-6.707887	.5981649
3 vs 2	1.278346	.6990888	1.83	0.405	-.5660303	3.122722
4 vs 2	-2.717142	.7341573	-3.70	0.001	-4.654038	-.7802462
4 vs 3	-3.995488	.7280521	-5.49	0.000	-5.916277	-2.074699

1._at:	2._at:	3._at:	4._at:
Low Green, Low Gini	Low Green, High Gini	High Green, Low Gini	High Green, High Gini

Table 12. predictive margins for three-way model.

Net immigration at:	*Green Service area at:*		*Margin*	*Delta-method std. err.*	*z*	*P>|z|*	*[95% conf.*	*interval]*
-0.05	0.05	High Gini	12.396	0.120	102.910	0.00	12.160	12.632
-0.05	0.25	High Gini	12.576	0.083	151.010	0.00	12.413	12.739
-0.05	0.45	High Gini	12.755	0.053	241.010	0.00	12.652	12.859
-0.05	0.65	High Gini	12.935	0.046	281.820	0.00	12.845	13.025
-0.05	0.85	High Gini	13.115	0.070	188.270	0.00	12.978	13.251
-0.05	0.05	Low Gini	12.383	0.179	69.250	0.00	12.032	12.733
-0.05	0.25	Low Gini	12.359	0.119	104.120	0.00	12.127	12.592
-0.05	0.45	Low Gini	12.336	0.064	191.870	0.00	12.210	12.462
-0.05	0.65	Low Gini	12.313	0.046	268.390	0.00	12.223	12.402
-0.05	0.85	Low Gini	12.289	0.089	137.360	0.00	12.114	12.464
0	0.05	High Gini	12.589	0.073	172.470	0.00	12.446	12.732
0	0.25	High Gini	12.684	0.046	272.990	0.00	12.593	12.775
0	0.45	High Gini	12.778	0.026	490.460	0.00	12.727	12.829
0	0.65	High Gini	12.873	0.030	431.870	0.00	12.815	12.931
0	0.85	High Gini	12.967	0.053	245.600	0.00	12.864	13.071
0	0.05	Low Gini	12.409	0.063	196.180	0.00	12.285	12.533
0	0.25	Low Gini	12.415	0.042	296.420	0.00	12.333	12.497
0	0.45	Low Gini	12.421	0.025	498.570	0.00	12.372	12.470
0	0.65	Low Gini	12.427	0.025	507.060	0.00	12.379	12.475
0	0.85	Low Gini	12.433	0.041	302.070	0.00	12.353	12.514
0.05	0.05	High Gini	12.782	0.083	154.610	0.00	12.620	12.945
0.05	0.25	High Gini	12.792	0.052	245.150	0.00	12.690	12.894
0.05	0.45	High Gini	12.801	0.026	497.120	0.00	12.751	12.852
0.05	0.65	High Gini	12.811	0.026	500.010	0.00	12.761	12.861
0.05	0.85	High Gini	12.820	0.052	246.600	0.00	12.718	12.922
0.05	0.05	Low Gini	12.436	0.116	107.440	0.00	12.209	12.662
0.05	0.25	Low Gini	12.471	0.074	168.300	0.00	12.326	12.616
0.05	0.45	Low Gini	12.506	0.040	308.910	0.00	12.427	12.586
0.05	0.65	Low Gini	12.542	0.043	292.270	0.00	12.458	12.626
0.05	0.85	Low Gini	12.577	0.078	161.070	0.00	12.424	12.730
0.1	0.05	High Gini	12.976	0.138	94.060	0.00	12.705	13.246
0.1	0.25	High Gini	12.900	0.093	138.890	0.00	12.718	13.082
0.1	0.45	High Gini	12.824	0.052	244.380	0.00	12.722	12.927
0.1	0.65	High Gini	12.749	0.038	339.450	0.00	12.675	12.822
0.1	0.85	High Gini	12.673	0.068	186.910	0.00	12.540	12.806
0.1	0.05	Low Gini	12.462	0.245	50.810	0.00	11.981	12.943
0.1	0.25	Low Gini	12.527	0.159	78.750	0.00	12.215	12.839
0.1	0.45	Low Gini	12.592	0.085	148.510	0.00	12.425	12.758
0.1	0.65	Low Gini	12.657	0.076	165.810	0.00	12.507	12.806
0.1	0.85	Low Gini	12.721	0.146	87.340	0.00	12.436	13.007

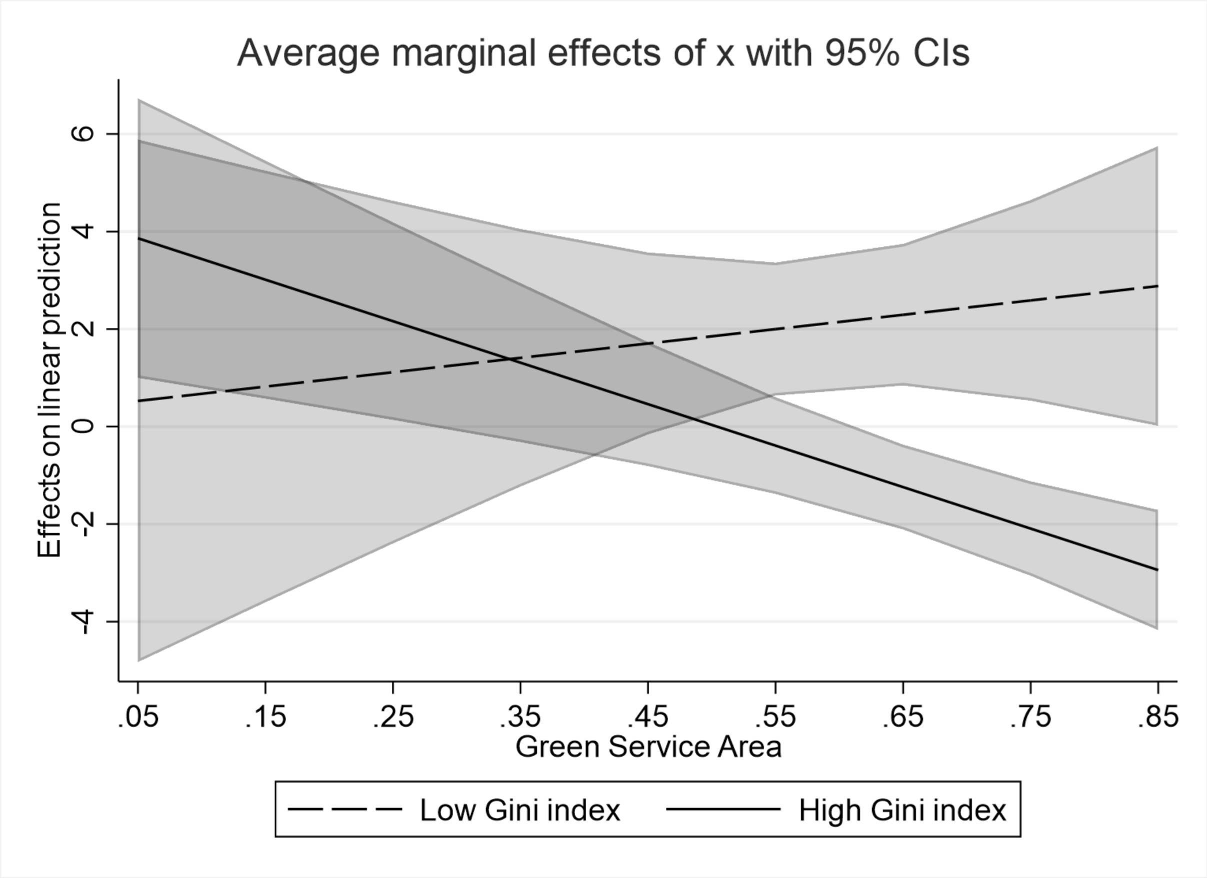


Figure 12. Marginsplot: Average marginal effects of the three-way moderation

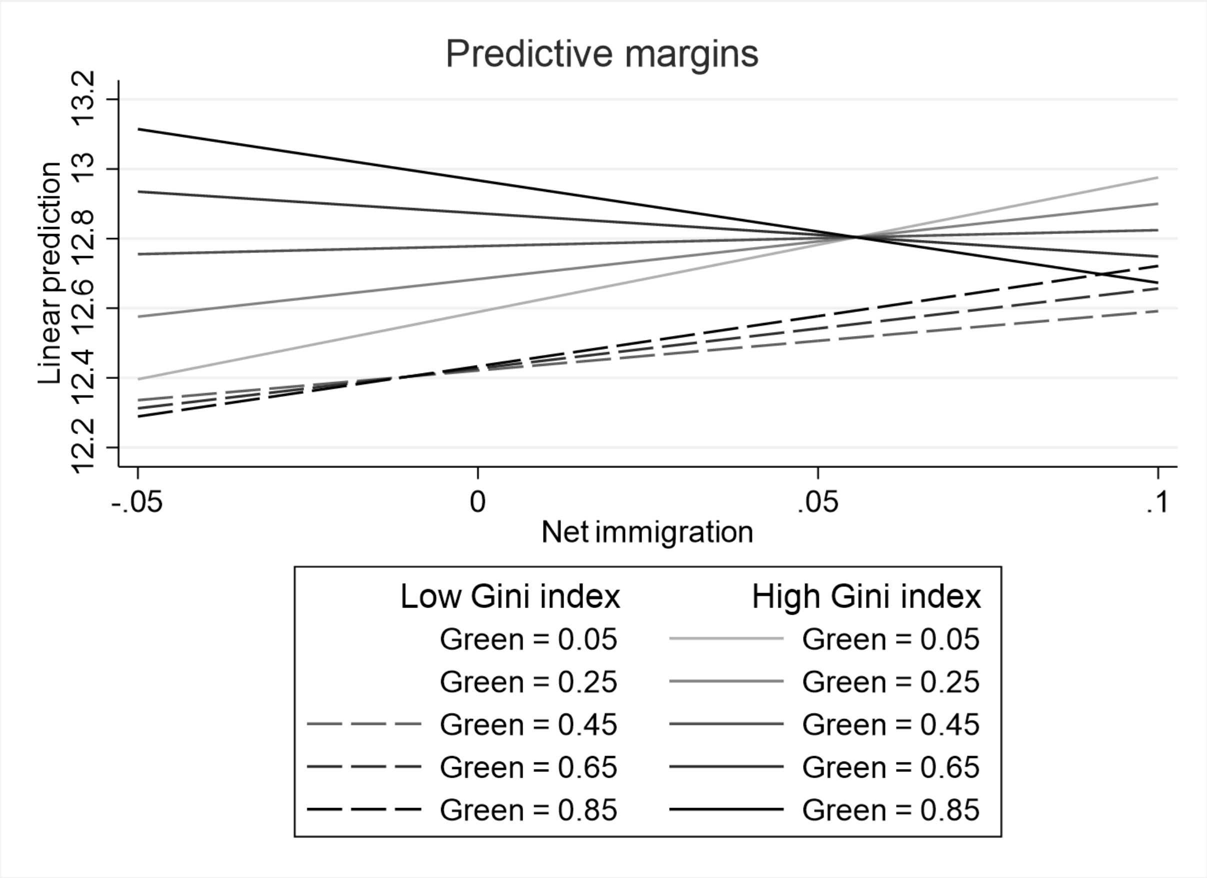


Figure 11. Marginsplot: Predictive margins of the three-way moderation.

At high Gini index the average rate of change in the response of log wage to one unit change in immigration increases by ca. 6.8 over increasing levels of green service area moderation (between individual neighborhoods over time). At low Gini the average rate of change in the response of log wage to one unit change in immigration decreases by ca. 2.3 over increasing levels of green service area moderation (between individual neighborhoods over time). Predictive margins may be able to illustrate the magnitude more clearly on the predicted values: At high Gini index negative immigration yields ca. 5.8% increase and positive immigration yields a ca. -2.4% change in predicted values over the green service area region of significance. At low Gini index negative immigration yields a ca. -0.4% change and at positive immigration a ca. 1.0% increase in the predicted values of log wage. Thus, the magnitude of the moderation at high income inequality is more notable than at low income inequality.

Next, comparing the pairwise differences in the slopes informs whether there is any significant difference in the main moderation based on the criteria of Gini index values. Plotting all the values of green service area yields too many slopes to meaningfully test pairwise differences, thus for this test one standard deviation above and below the mean is used for both moderators:

. margins, dydx(*net immigration*) at(*green service area*=(0.05(0.1)0.85) *low gini, high gini)* pwcompare(effects) mcpmpare(Bonferroni)

The output seen in Table 11 shows that the moderating effect of high green at high Gini is statistically significant from low green at high Gini. Also, high green at high Gini is significantly different from high green at low Gini. The differences are evident in the graphs, and supports an interpretation of the moderating effect of Gini index on the main moderation. Interpretating the predicted margins in the output and the graphs suggests that the moderating effect of green service area on the relation between net immigration and log wage is itself moderated by the level of gini index. Under conditions of high Gini index the moderating effect of green service area at higher value is largely unchanged. However, at high Gini index values the moderating effect of lower levels of green service area changes direction, still having a negative impact on the main effect at negative values of net immigration, but now show a positive impact on the main effect. Also, at lower Gini index levels the moderating effect of higher levels of green service area similarly has a positive impact on the main effect at high positive values of net immigration, while having a negative impact at negative net migration. Hence, the casual interpretation is that: **High Gini, high green service area:** In neighborhoods with high income inequality increasing access to green space in 300 meters positively affects the migration-wage relation at negative immigration rates (suggesting a possible displacement of lower income residents), while negatively affecting the immigration-wage relation at higher positive immigration rates (suggesting lower immigration rate of higher income residents). **High Gini, low green service area:** In neighborhoods with high income inequality lower access to green space positively affects the relation between immigration and wage at higher positive immigration rates (suggesting an increasing immigration of higher income residents) and negatively affects the main relation at negative migration rates (suggesting a possible emigration of higher income residents, yet not necessarily displacing lower income residents). **Low Gini, medium to high green service area:** In neighborhoods with low income inequality increasing access to green space is overall associated with higher wages as migration increase (suggesting that the low income inequality may have few poor residents that can be displaced). These findings seem to support the hypothesized moderation effect on socioeconomic mobility.

5.2 Discussion

Based on the literature it was hypothesized that higher coverage of access to green spaces increases the average income level of a neighborhood by attracting higher income residents, while simultaneously in areas with higher income inequality this population mobility could foster gentrification by displacing lower income residents, thus also increasing wage levels. Specifically, based on the literature the aim of the research was to investigate whether these effects are evident for medium to large green spaces within a distance parameter of 300 meters.

The two-way interaction supports the finding that overall increased access to green spaces has a positive impact on attracting increasing income levels through migration, and especially so at negative migration levels, which suggests that displacement of poorer residents does occur. Furthermore, the three-way interaction supports the hypothesized mechanism and elaborates on the findings by specifying how the effect of access to green space varies across levels of income inequality. It finds when increasing the level of access to green space in a neighborhood the gentrification effect is mostly evident in areas with high income inequality, while in areas with lower income inequality the hedonic pricing effect shows less evidence of gentrification. Additionally, both models provide rough estimates of the predicted magnitudes of moderation. The findings should be of interest to historic as well as current green infrastructure policies of Copenhagen municipality aiming toward improvements of agglomeration economics through urban environmental sustainability, while having to balance these changes with measures to ensure socially responsible outcomes. The results of the analysis may help narrow down the parameters and conditions where green infrastructure policies can achieve their desired effects.

Of note, the analysis is built on the assumption that the relationship between net immigration and wage translates into socioeconomic mobility. While this outcome is likely, given the direct association between the two variables the connection to hedonic pricing and gentrification may not necessarily be the true relation between the two variables, since causes other than those two may cause the wage response to migration. Additionally, the moderating effect of Gini index levels only informs the responses on wage from the main moderation, but does not directly inform whether the area has a become more or less unequal in wages resulting from migration. The variable for access to green space is based on geographic designations of neighborhoods, which as was mentioned in Chapter 2, may be subject to geographic selection bias from the "modifiable areal unit problem," as the parameters of the analysis remains untested for accurate cross-tabulation to ensure stable estimates at different geographic scales and parameters. Additionally, the data for green spaces is time lagged. While the effect is assumed to be accumulative, it may limit the preciseness when estimating the effects of access to green space. Finally, the analysis is based on ecological inference of aggregated data, which does not necessarily reflect the true behavior of individuals. Therefore, assuming that the fixed effects estimator allows for causal inference of between-unit effects for the interaction effects should be done so only with some confidence, which could be improved with individual level data. With those limitations in mind, the findings from the analysis should allow careful inference to the interest of policy making, and provides a good basis for further research on the topic, including different geographic and temporal parameters, analysis of individual-level data, and different specifications to better measure responses, e.g. housing prices and income inequality.

6 Conclusion

6.1 Research Aims

The aim of Copenhagen municipality to attract socio-economically strong residents with its green infrastructure strategy inspired the analysis which aimed to assess whether improving access green areas in the city has had any impact on socio-economic mobility of its residents. Additionally, the concerns of potential displacement of socio-economically weaker residents warranted consideration of how income inequality affects the impact of access to green spaces. Based on the literature it was hypothesized that higher coverage of access to green spaces increases the average income level of a neighborhood by attracting higher income residents, while simultaneously in areas with higher income inequality this population mobility could foster gentrification by displacing lower income residents, thus also increasing wage levels.

6.2 Research Objectives

As such, the objective of the research was to investigate how access to green spaces within 300 meters is related to socioeconomic migration in Copenhagen, specifically whether the development of access to green spaces in Copenhagen municipality significantly attracted higher income residents, and if this effect have simultaneously displaced low-income residents. The geographical scope of the analysis covers the 67 distinct neighborhoods in Copenhagen, while the temporal scope of the analysis covers the interval of years 2008, 2012 and 2018. In order to investigate the hypothesized mechanism of access to green space increasing neighborhood income levels by attracting higher-income residents, and potentially displacing lower-income residents, it is first necessary to investigate whether there is an independent interaction in the first relationship between access to green space, migration, and income levels. Thereafter, the second interaction is investigated by including the effect of income inequality. Thus, a fixed effects regression analysis of a two-way and three-way interaction was conducted.

6.3 Practical Implications

The two-way interaction supported the finding that overall increased access to green spaces has a positive impact on attracting increasing income levels through migration, and especially so at negative migration levels, which suggested that displacement of poorer residents occurs. Furthermore, the three-way interaction supported the hypothesized mechanism and specified how the effect of access to green space varies for areas with different income inequality levels. It found that in increasing levels of access to green space in a neighborhood the gentrification effect is mostly evident in areas with high income inequality, while in areas with lower income inequality the hedonic pricing effect shows less evidence of gentrification. The findings should be of interest to historic as well as current green infrastructure policies of Copenhagen municipality aiming toward improvements of agglomeration economics through urban environmental sustainability, while having to balance these changes with measures to ensure socially responsible outcomes. The results of the analysis may help narrow down the parameters and conditions where green infrastructure policies can achieve their desired effects.

6.4 Future Research

The findings from the analysis should allow careful inference to the interest of policy making, which provides a good basis for further research on the topic, and may include areas such as: Verification of estimates at different geographic and temporal parameters; Analysis of individual-level data for improved causal inference; Different specifications to better measure responses of the interaction effects, such as responses of housing prices and income inequality.

6.5 Chapter Summary

Chapter 1 covers the introduction to the green infrastructure policies of Copenhagen, where increasing access to green space is a key to its environmental and socioeconomic objectives. Chapter 2 investigates the literature review of theory on the effects of access to green space on socioeconomic migration, from which the hypothesized interaction models are derived. Chapter 3 overviews the key variables of wage, net immigration, green service area, gini index, as well as control variables, and considers the ecological inference of aggregated panel data. Chapter 4 outlines the methodology used to conduct the analysis, namely analysis of margins in fixed effects regressions with two-way and three-way interactions between the key variables. Chapter 5 examines and the results of the average marginal effects and predictive margins, and discusses the validity of the findings, which ultimately support the hypothesized relationships.

References

Abendroth, A.-K., Huffman, M. L. & Treas, J. (2014). The Parity Penalty in Life Course Perspective: Motherhood and Occupational Status in 13 European Countries, *American Sociological Review*, vol. 79, no. 5, pp.993–1014.

Agency of Data Supply and Infrastructure. (2022). The Danish Administrative Geographical Divison (DAGI), Available Online: http://eng.sdfe.dk/products-and-services/the-danish-administrative-geographical-division-dagi [Accessed 21 July 2022].

Aiken, L. S. & West, S. G. (1991). Multiple Regression: Testing and Interpreting Interactions, Thousand Oaks, CA, US: Sage Publications, Inc, pp.xi, 212.

Allison, P. D. (1977). Testing for Interaction in Multiple Regression, *American Journal of Sociology*, vol. 83, no. 1, pp.144–153.

Allison, P. D. (2009). Fixed Effects Regression Models, SAGE Publications.

Ananat, E., Shihe, F. & Ross, S. L. (2018). Race-Specific Urban Wage Premia and the Black-White Wage Gap, *Journal of Urban Economics*, vol. 108, pp.141–153.

Andreß, H.-J., Golsch, K. & Schmidt, A. W. (2013). Applied Panel Data Analysis for Economic and Social Surveys, Springer Berlin Heidelberg.

Anguelovski, I., Connolly, J. J. T., Masip, L. & Pearsall, H. (2018). Assessing Green Gentrification in Historically Disenfranchised Neighborhoods: A Longitudinal and Spatial Analysis of Barcelona, *Urban Geography*, vol. 39, no. 3, pp.458–491.

Baltagi, B. H. (2005). Econometric Analysis of Panel Data, Wiley.

Barbosa, O., Tratalos, J. A., Armsworth, P. R., Davies, R. G., Fuller, R. A., Johnson, P. & Gaston, K. J. (2007). Who Benefits from Access to Green Space? A Case Study from Sheffield, UK, *Landscape and Urban Planning*, vol. 83, no. 2, pp.187–195.

Bauer, D. J. & Curran, P. J. (2005). Probing Interactions in Fixed and Multilevel Regression: Inferential and Graphical Techniques, *Multivariate Behavioral Research*, vol. 40, no. 3, pp.373–400.

Bell, A., Fairbrother, M. & Jones, K. (2019). Fixed and Random Effects Models: Making an Informed Choice, *Quality & Quantity*, vol. 53, no. 2, pp.1051–1074.

Bijker, R. A. & Sijtsma, F. J. (2017). A Portfolio of Natural Places: Using a Participatory GIS Tool to Compare the Appreciation and Use of Green Spaces inside and Outside Urban Areas by Urban Residents, *Landscape and Urban Planning*, vol. 158, pp.155–165.

Born, B. & Breitung, J. (2016). Testing for Serial Correlation in Fixed-Effects Panel Data Models, *Econometric Reviews*, vol. 35, no. 7, pp.1290–1316.

Brüderl, J. & Ludwig, V. (2014). Fixed-Effects Panel Regression, in *The SAGE Handbook of Regression Analysis and Causal Inference*, [e-book] 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, pp.327–358, Available Online: https://methods.sagepub.com/book/regression-analysis-and-causal-inference/n15.xml [Accessed 2 August 2022].

Cameron, A. C., Trivedi, P. K. & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications, Cambridge University Press.

Cameron, A. & Trivedi, P. (2010). Microeconometrics Using Stata, Revised Edition, Stata Press books, StataCorp LP, Available Online: https://econpapers.repec.org/bookchap/tsjspbook/musr.htm [Accessed 2 August 2022].

Center for Urban Development. (2014). Bosætning i København En analyse af bosætningsmønstre og boligpræferencer, Copenhagen Municipality, The Finance Committee, p.62.

Cheshire, P. & Sheppard, S. (2005). The Introduction of Price Signals into Land Use Planning Decision-Making: A Proposal, *Urban Studies*, vol. 42, no. 4, pp.647–663.

City of Copenhagen: Statbank - Data and Statistics. (2022). , Available Online: https://kk.statistikbank.dk/statbank5a/SelectTable/Omrade0.asp?SubjectCode=&PLanguage=1 [Accessed 18 July 2022].

Cömertler, S. (2017). Greens of the European Green Capitals, *IOP Conference Series: Materials Science and Engineering*, vol. 245, p.052064.

Connolly, J. J. T. (2019). From Jacobs to the Just City: A Foundation for Challenging the Green Planning Orthodoxy, vol. 91, pp.64–70.

Copenhagen Municipality, C. for E. (2022). E-Mail Correspondance: Neighborhood Divisions.

CORINE Land Cover — Copernicus Land Monitoring Service. (2022). *Copernicus Land Monitoring Service*, Available Online: https://land.copernicus.eu/pan-european/corine-land-cover [Accessed 29 April 2022].

Creswell, J. W. & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th edition., Los Angeles: SAGE Publications, Inc.

Crompton, J. L. (2001). The Impact of Parks on Property Values: A Review of the Empirical Evidence, *Journal of Leisure Research*, vol. 33, no. 1, pp.1–31.

Damm, A. P., Schultz-Nielsen, M. L. & Tranæs, A. P. D., Marie-Louise Schultz-Nielsen,Torben. (2016). En befolkning deler sig op?, Gyldendal A/S.

Dawson, J. F. (2014). Moderation in Management Research: What, Why, When, and How, *Journal of Business and Psychology*, vol. 29, no. 1, pp.1–19.

Dawson, J. & Richter, A. (2006). Probing Three-Way Interactions in Moderated Multiple Regression: Development and Application of a Slope Difference Test., *The Journal of applied psychology*.

Esri. (2022). ArcGIS Pro, *ArcGIS Pro*, Available Online: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview [Accessed 21 July 2022].

European Commision. (2022). European Green Capital Award, *European Green Capital Award*, Available Online: https://environment.ec.europa.eu/topics/urban-environment/european-green-capital-award_en [Accessed 18 August 2022].

Evangelio, R., Hone, S., Lee, M. & Prentice, D. (2019). What Makes a Locality Attractive? Estimates of the Amenity Value of Parks for Victoria, *Economic Papers: A journal of applied economics and policy*, vol. 38, no. 3, pp.182–192.

Farris, F. A. (2010). The Gini Index and Measures of Inequality, *The American Mathematical Monthly*, vol. 117, no. 10, pp.851–864.

Feng, S., Li, G., Tong, T. & Luo, S. (2020). Testing for Heteroskedasticity in Two-Way Fixed Effects Panel Data Models, *Journal of Applied Statistics*, vol. 47, no. 1, pp.91–116.

Finsaas, M. C. & Goldstein, B. L. (2021). Do Simple Slopes Follow-up Tests Lead Us Astray? Advancements in the Visualization and Reporting of Interactions, *Psychological Methods*, vol. 26, no. 1, pp.38–60.

Firebaugh, G., Warner, C. & Massoglia, M. (2013). Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis, 2013.

Gangl, M. (2010). Causal Inference in Sociological Research, *Annual Review of Sociology*, vol. 36, no. 1, pp.21–47.

Gastwirth, J. L. (1972). The Estimation of the Lorenz Curve and Gini Index, *The Review of Economics and Statistics*, vol. 54, no. 3, pp.306–316.

Gehl, J., Kaefer, L. & Reigstad, S. (2006). Close Encounters with Buildings, *Urban Design International*, vol. 11, pp.29–47.

Giesselmann, M. & Schmidt-Catran, A. W. (2022). Interactions in Fixed Effects Regression Models, *Sociological Methods & Research*, vol. 51, no. 3, pp.1100–1127.

Goossens, C., Oosterlynck, S. & Bradt, L. (2020). Livable Streets? Green Gentrification and the Displacement of Longtime Residents in Ghent, Belgium, *Urban Geography*, vol. 41, no. 4, pp.550–572.

Government of Denmark. (2019). Danmarks hovedstad - Initiativer til styrkelse af hovedstadsområdet, p.44, Available Online: https://em.dk/media/12899/hovedstaden-2030_web_accessible.pdf.

Green Infrastructure. (2022). , Available Online: https://biodiversity.europa.eu/countries/denmark/green-infrastructure [Accessed 20 July 2022].

Haase, D., Kabisch, S., Haase, A., Andersson, E., Banzhaf, E., Baró, F., Brenck, M., Fischer, L. K., Frantzeskaki, N., Kabisch, N., Krellenberg, K., Kremer, P., Kronenberg, J., Larondelle, N., Mathey, J., Pauleit, S., Ring, I., Rink, D., Schwarz, N. & Wolff, M. (2017). Greening Cities – To Be Socially Inclusive? About the Alleged Paradox of Society and Ecology in Cities, *Habitat International*, vol. 64, pp.41–48.

Halaby, C. N. (2004). Panel Models in Sociological Research: Theory into Practice, *Annual Review of Sociology*, vol. 30, pp.507–544.

Hausman, J. A. (1978). Specification Tests in Econometrics, *Econometrica*, vol. 46, no. 6, pp.1251–1271.

Hayes, A. F. (2022). Introduction to Mediation, Moderation, and Conditional Process Analysis, Third Edition: A Regression-Based Approach, Guilford Publications.

Heward, J. (2022). Copenhagen - the Green City, *Danish Architecture and Design Review*, Available Online: http://danishdesignreview.com/kbhnotes/2020/1/18/copenhagen-urban-nature [Accessed 25 July 2022].

Holm, S. & Jakobsen, C. (2006). Rekreativ Brug Af Byens Grønne Områder, 31, Skov og Landskab, Available Online: https://videntjenesten.ku.dk/filer/rapporter/park-og-landskab/pogl31.pdf [Accessed 25 July 2022].

Huber, C. (2022). In the Spotlight: Visualizing Continuous-by-Continuous Interactions with Margins and Twoway Contour, *Www.Stata.Com*, Available Online: https://www.stata.com/stata-news/news32-1/spotlight/ [Accessed 4 August 2022].

Immergluck, D. & Balan, T. (2018). Sustainable for Whom? Green Urban Development, Environmental Gentrification, and the Atlanta Beltline, *Urban Geography*, vol. 39, no. 4, pp.546–562.

Inoue, A. & Solon, G. (2006). A Portmanteau Test for Serially Correlated Errors in Fixed Effects Models, *Econometric Theory*, vol. 22, pp.835–851.

Jaccard, J. & Turrisi, R. (2003). Interaction Effects in Multiple Regression, [e-book] 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc., Available Online: https://methods.sagepub.com/book/interaction-effects-in-multiple-regression [Accessed 2 August 2022].

Johnson, P. O. & Neyman, J. (1936). Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems, *Statistical Research Memoirs*, vol. 1, pp.57–93.

Kennedy, P. (2008). A Guide to Econometrics, Malden, MA: Blackwell Pub.

Killewald, A. & Gough, M. (2013). Does Specialization Explain Marriage Penalties and Premiums?, *American Sociological Review*, vol. 78, no. 3, pp.477–502.

Kim, S. K. & Wu, L. (2022). Do the Characteristics of New Green Space Contribute to Gentrification?, *Urban Studies*, vol. 59, no. 2, pp.360–380.

King, G. (1997). A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data, Princeton, NJ: Princeton University Press.

Koefoed, O. (2019). Urban Nature as Transformed Practice – A Case of Multi-Dimensional Processing to Increase Public Value in Copenhagen, *Local Economy: The Journal of the Local Economy Policy Unit*, vol. 34, pp.525–544.

Kong, F., Yin, H. & Nakagoshi, N. (2007). Using GIS and Landscape Metrics in the Hedonic Price Modeling of the Amenity Value of Urban Green Space: A Case Study in Jinan City, China, *Landscape and Urban Planning*, vol. 79, no. 3, pp.240–252.

Kristensen, J. & Larsen, J. E. (2007). Fattigdom, social eksklusion og boligforhold, 4, *Dansk Sociologi*, vol. 18, no. 4, pp.9–28.

Kühhirt, M. (2012). Childbirth and the Long-Term Division of Labour within Couples: How Do Substitution, Bargaining Power, and Norms Affect Parents' Time Allocation in West Germany?, *European Sociological Review*, vol. 28, no. 5, pp.565–582.

Lemieux, T. (2006). The "Mincer Equation" Thirty Years After Schooling, Experience, and Earnings, in S. Grossbard (ed.), *Jacob Mincer A Pioneer of Modern Labor Economics*, [e-book] Boston, MA: Springer US, pp.127–145, Available Online: https://doi.org/10.1007/0-387-29175-X_11 [Accessed 20 July 2022].

Luttik, J. (2000). The Value of Trees, Water and Open Space as Reflected by House Prices in the Netherlands, *Landscape and Urban Planning*, vol. 48, no. 3, pp.161–167.

Maas, J., Verheij, R., Groenewegen, P., de Vries, S. & Spreeuwenberg, P. (2006). Green Space, Urbanity, and Health: How Strong Is the Relation?, *Journal of epidemiology and community health*, vol. 60, pp.587–92.

Masayuki, M. (n.d.). Urban Density, Human Capital, and Productivity: An Empirical Analysis Using Wage Data, p.28.

Mincer, J. A. (1974). Schooling, Experience, and Earnings, Available Online: https://www.nber.org/books-and-chapters/schooling-experience-and-earnings [Accessed 20 July 2022].

Morgan, S. L. & Winship, C. (2014). Counterfactuals and Causal Inference: Methods and Principles for Social Research, 2nd edn, [e-book] Cambridge: Cambridge University Press, Available Online: https://www.cambridge.org/core/books/counterfactuals-and-causal-inference/5CC81E6DF63C5E5A8B88F79D45E1D1B7 [Accessed 2 August 2022].

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data, *Econometrica*, vol. 46, no. 1, pp.69–85.

Nielsen, T. S. & Hansen, K. B. (2006). Nearby Nature and Green Areas Encourage Outdoor Activitiesand Decrease Mental Stress., *CAB Reviews*, [e-journal], Available Online: https://doi.org/10.1079/PAVSNNR20061059 [Accessed 25 July 2022].

Nordh, H. & Olafsson, A. S. (2021). Plans for Urban Green Infrastructure in Scandinavia, *Journal of Environmental Planning and Management*, vol. 64, no. 5, pp.883–904.

Oesch, D. & Lipps, O. (2013). Does Unemployment Hurt Less If There Is More of It Around? A Panel Analysis of Life Satisfaction in Germany and Switzerland, *European Sociological Review*, vol. 29, no. 5, pp.955–967.

Plümper, T., Troeger, V. E. & Manow, P. (2005). Panel Data Analysis in Comparative Politics: Linking Method to Theory, *European Journal of Political Research*, vol. 44, no. 2, pp.327–354.

Politiken. (2019). Alle københavnere skal have et stykke med grønt inden for 300 meter fra hjemmet, *Politiken*, 26 November, Available Online: https://politiken.dk/indland/kobenhavn/art7517882/Alle-k%C3%B8benhavnere-skal-have-et-stykke-med-gr%C3%B8nt-inden-for-300-meter-fra-hjemmet [Accessed 28 July 2022].

Preacher, K. J., Curran, P. J. & Bauer, D. J. (2006). Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent Curve Analysis, *Journal of Educational and Behavioral Statistics*, vol. 31, no. 4, pp.437–448.

Preacher, K. J., Rucker, D. D. & Hayes, A. F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions, *Multivariate Behavioral Research*, vol. 42, no. 1, pp.185–227.

Quality in the Production of Statistics. (2022). , Available Online: https://www.dst.dk/en/OmDS/strategi-og-kvalitet/kvalitet-for-statistikproduktion [Accessed 18 July 2022].

Razykova, V., Bartke, S. & Schwarz, N. (2014). A Dynamic Model of Urban Green Spaces, Residential Mobility and Real Estate Market, *Advances in Computational Social Science and Social Simulation Barcelona: Autònoma University of Barcelona*, p.2.

Rigolon, A. & Nemeth, J. (2018). “We’re Not in the Business of Housing:” Environmental Gentrification and the Nonprofitization of Green Infrastructure Projects, *Cities*.

Rigolon, A. & Németh, J. (2020). Green Gentrification or 'Just Green Enough': Do Park Location, Size and Function Affect Whether a Place Gentrifies or Not?, *Urban Studies*, vol. 57, no. 2, pp.402–420.

Rising, B. (2011). Graphics Tricks for Models, Graphics Tricks for Models, 15 2011, p.25, Available Online: https://www.stata.com/meeting/chicago11/materials/chi11_rising.pdf.

Rogosa, D. (1980). Comparing Nonparallel Regression Lines, *Psychological Bulletin*, vol. 88, no. 2, pp.307–321.

Rosenbak, M. & Jørgensen, G. (2009). Den Gønne by - Udfordringer Og Muligheder, *Naturstyrelsen*, Available Online: https://naturstyrelsen.dk/media/nst/attachments/den_groenne_by_sl.pdf [Accessed 21 July 2022].

Saeed, A., Aslam, M., Altaf, S. & Amanullah, M. (2018). Improved Inference for the Panel Data Model with Unknown Unit-Specific Heteroscedasticity: A Monte Carlo Evidence, *Cogent Mathematics & Statistics*, vol. 5, no. 1, p.1463598.

Schipperijn, J., Ekholm, O., Stigsdotter, U. K., Toftager, M., Bentsen, P., Kamper-Jørgensen, F. & Randrup, T. B. (2010). Factors Influencing the Use of Green Space: Results from a Danish National Representative Survey, *Landscape and Urban Planning*, vol. 95, no. 3, pp.130–137.

Schofer, E. & Longhofer, W. (2011). The Structural Sources of Association, *American Journal of Sociology*, vol. 117, no. 2, pp.539–585.

Schunck, R. (2013). Within and between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models, *The Stata Journal*, vol. 13, no. 1, pp.65–76.

Schwarz, N., Haase, A., Haase, D., Kabisch, N., Kabisch, S., Liebelt, V., Rink, D., Strohbach, M. W., Welz, J. & Wolff, M. (2021). How Are Urban Green Spaces and Residential Development Related? A Synopsis of Multi-Perspective Analyses for Leipzig, Germany, 6, *Land*, vol. 10, no. 6, p.630.

Sharifi, F., Nygaard, A., Stone, W. M. & Levin, I. (2021). Accessing Green Space in Melbourne: Measuring Inequity and Household Mobility, *Landscape and Urban Planning*, vol. 207, p.104004.

Shaver, J. M. (2019). Interpreting Interactions in Linear Fixed-Effect Regression Models: When Fixed-Effect Estimates Are No Longer Within-Effects, *Strategy Science*, vol. 4, no. 1, pp.25–40.

Sheppard, S. & Cheshire, P. (1995). On the Price of Land and the Value of Amenities, *Economica*, vol. 62, pp.247–67.

Skirbekk, V. (2004). Age and Individual Productivity, *Vienna Yearbook of Population Research*.

Stata. (2021). Stata Base Reference Manual, Release 17, [e-book] Stata Press, Available Online: https://www.stata.com/bookstore/base-reference-manual/ [Accessed 5 August 2022].

StatBank Denmark. (2022). , Available Online: https://www.statbank.dk/statbank5a/default.asp?w=1368 [Accessed 18 July 2022].

Steinhardt, M. F. (2012). Does Citizenship Matter? The Economic Impact of Naturalizations in Germany, *Labour Economics*, vol. 19, no. 6, pp.813–823.

Stigsdotter, U. K., Ekholm, O., Schipperijn, J., Toftager, M., Kamper-Jørgensen, F. & Randrup, T. B. (2010). Health Promoting Outdoor Environments - Associations between Green Space, and Health, Health-Related Quality of Life and Stress Based on a Danish National Representative Survey, *Scandinavian Journal of Public Health*, vol. 38, no. 4, pp.411–417.

Sullivan, W. C., Kuo, F. E. & Depooter, S. F. (2004). The Fruit of Urban Nature: Vital Neighborhood Spaces, *Environment and Behavior*, vol. 36, no. 5, pp.678–700.

The Technical and Environmental Administration. (2015). Urban Nature in Copenhagen - Strategy 2015-2025 2015, Copenhagen Municipality, Available Online: https://kk.sites.itera.dk/apps/kk_pub2/index.asp?id=1653&mode=detalje [Accessed 27 July 2022].

The Technical and Environmental Administration. (2018). Analyse Af Behov for Store Og Melemstore Grønne Områder i København - Samt Analyse Af Potentialer for at Udvikle Københavns Grønne Og Blå Struktur, The Technical and Environmental Administration, Available Online: https://kp19.kk.dk/sites/default/files/2021-03/analyse_af_behov_for_store_og_mellemstore_groenne_omraader_14._februar_compressed.pdf [Accessed 5 April 2022].

UCLA. (2022a). How Can I Explain a Continuous by Continuous Interaction?, Available Online: https://stats.oarc.ucla.edu/stata/faq/how-can-i-explain-a-continuous-by-continuous-interaction-stata-12/ [Accessed 14 July 2022].

UCLA. (2022b). How Can I Understand a 3-Way Continuous Interaction?, Available Online: https://stats.oarc.ucla.edu/stata/faq/how-can-i-understand-a-3-way-continuous-interaction-stata-12/ [Accessed 11 July 2022].

Urban Atlas — Copernicus Land Monitoring Service. (2022). *Copernicus Land Monitoring Service*, Available Online: https://land.copernicus.eu/local/urban-atlas [Accessed 1 May 2022].

Wheeler, C. H. (2004). Wage Inequality and Urban Density, *Journal of Economic Geography*, vol. 4, no. 4, pp.421–437.

Williams, R. (2012). Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects, *The Stata Journal*, vol. 12, no. 2, pp.308–331.

Wolch, J. R., Byrne, J. & Newell, J. P. (2014). Urban Green Space, Public Health, and Environmental Justice: The Challenge of Making Cities 'Just Green Enough', *Landscape and Urban Planning*, vol. 125, pp.234–244.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data, [e-book] The MIT Press, Available Online: https://www.jstor.org/stable/j.ctt5hhcfr [Accessed 2 August 2022].

Wu, J. & Plantinga, A. J. (2003). The Influence of Public Open Space on Urban Spatial Structure, *Journal of Environmental Economics and Management*, vol. 46, no. 2, pp.288–309.

Young, R. & Johnson, D. R. (2015). Handling Missing Values in Longitudinal Panel Data With Multiple Imputation, *Journal of marriage and the family*, vol. 77, no. 1, pp.277–294.

Zveglich Jr., J., Rodgers, Y. van der M. & Lavina, E. (2019). Expected Work Experience:: A New Human Capital Measure, ADB Economics Working Paper Series, 0 edn, Manila, Philippines: Asian Development Bank, Available Online: https://www.adb.org/publications/expected-work-experience-new-human-capital-measure [Accessed 23 July 2022].

Appendix A

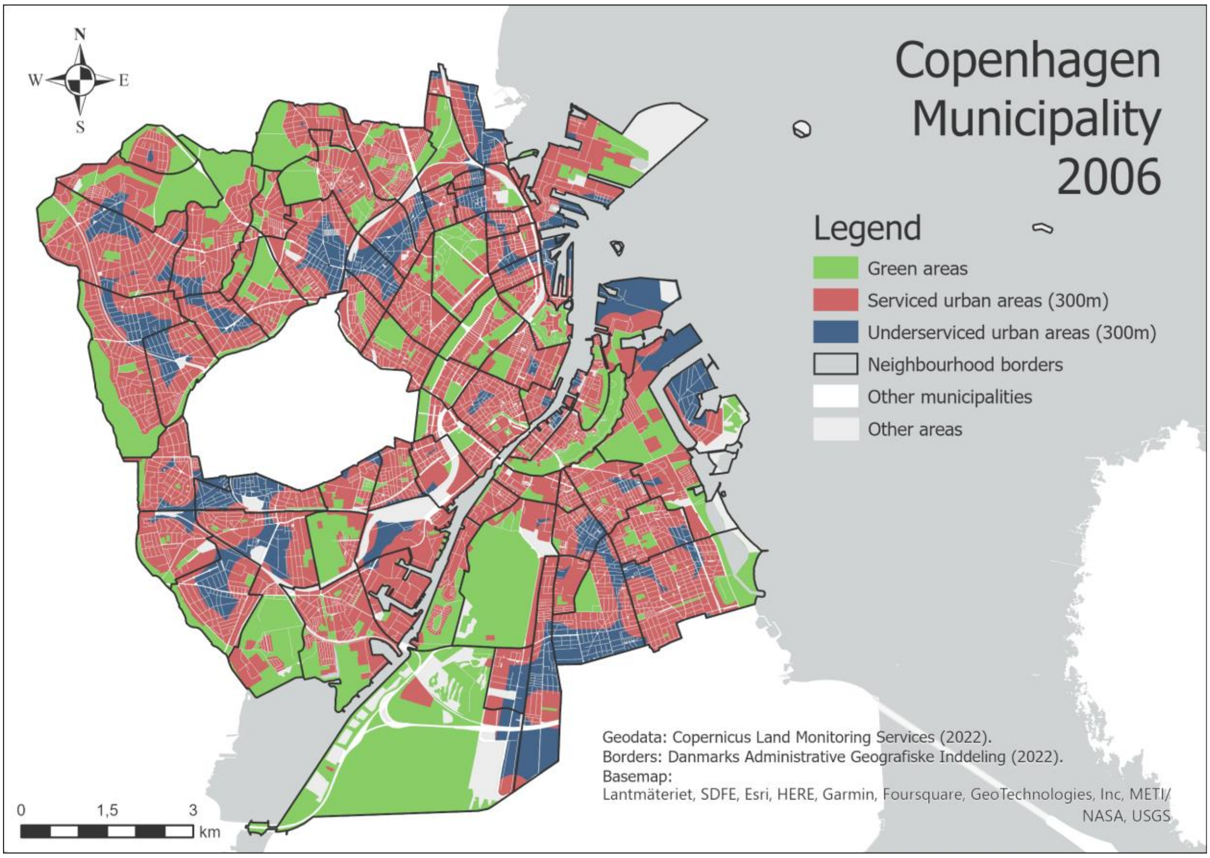


Figure 14. Green space 300m service area (2006).

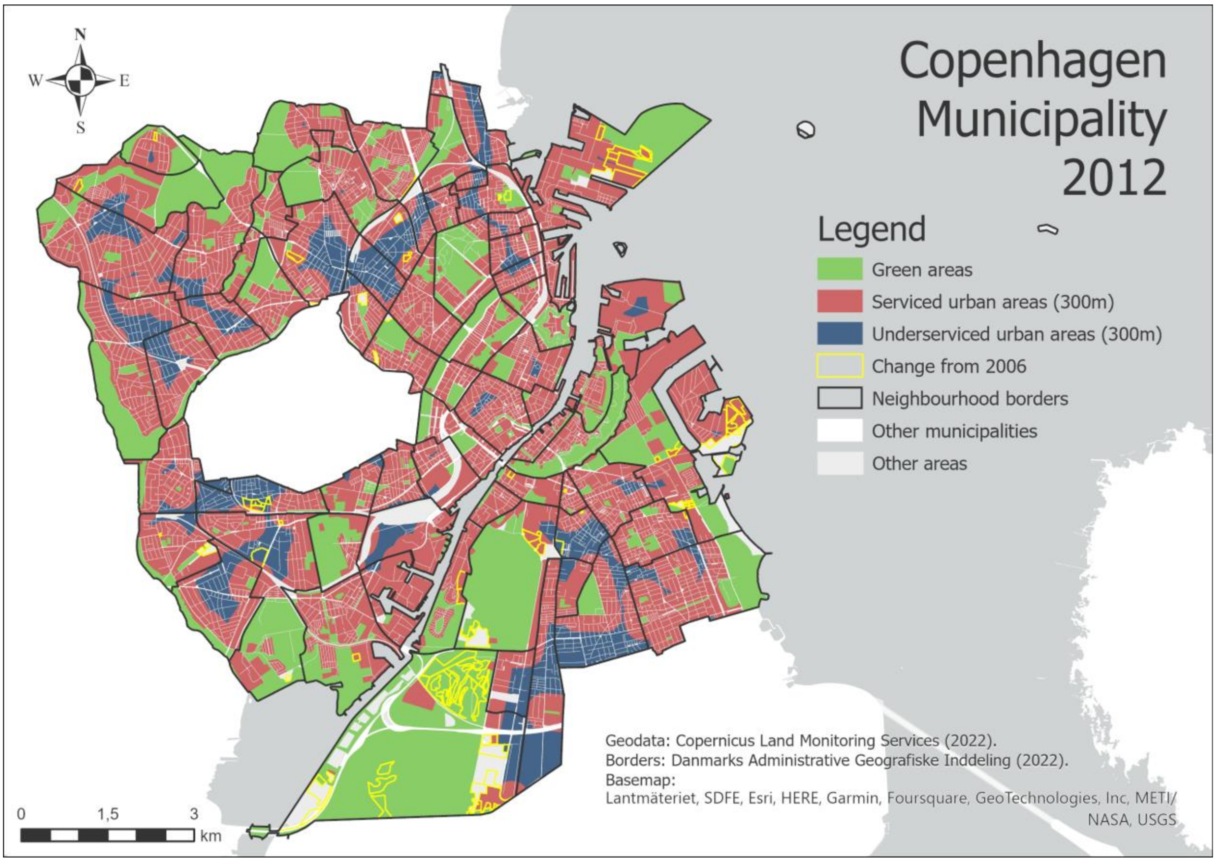


Figure 13. Green space 300m service area (2012)

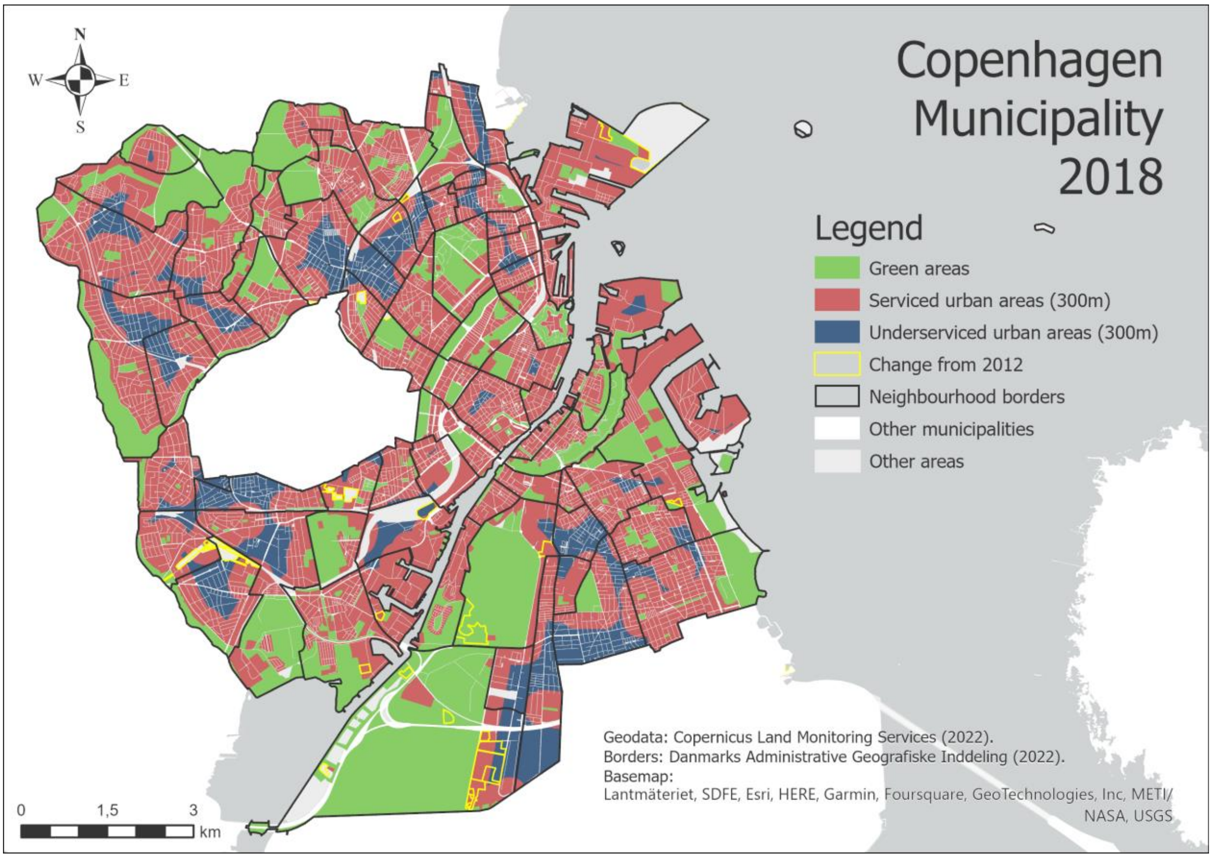


Figure C3. Green space 300m service area (2018)

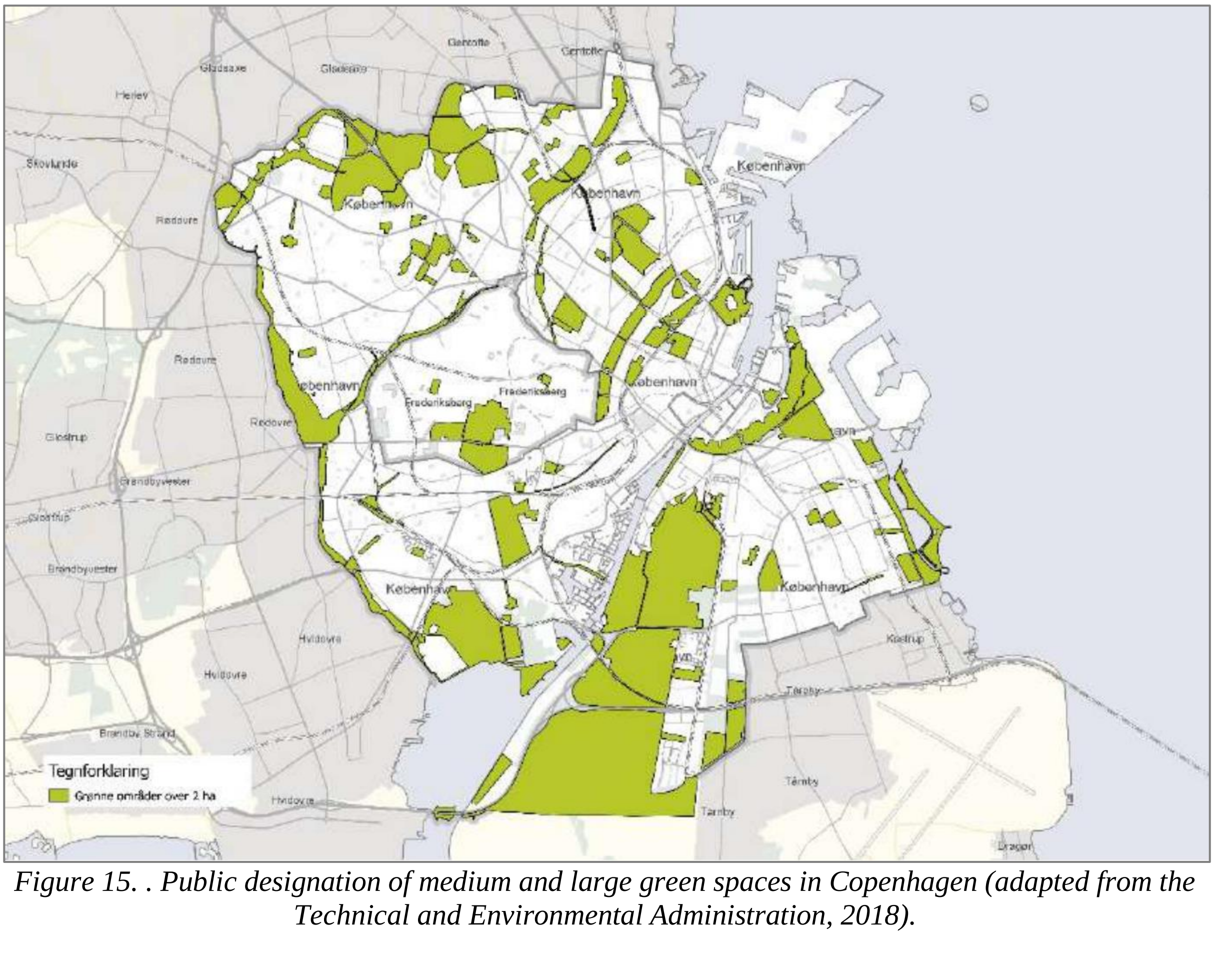


Figure 15. . Public designation of medium and large green spaces in Copenhagen (adapted from the Technical and Environmental Administration, 2018).

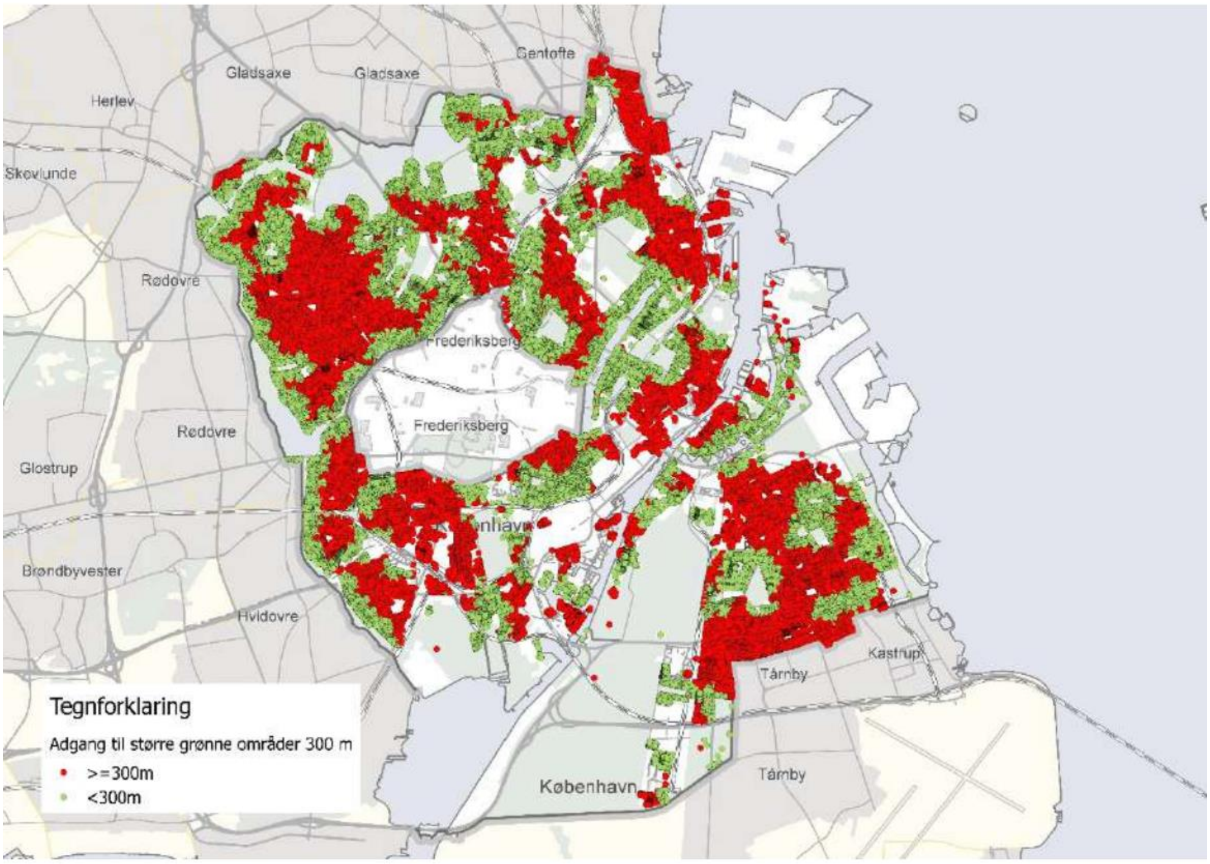
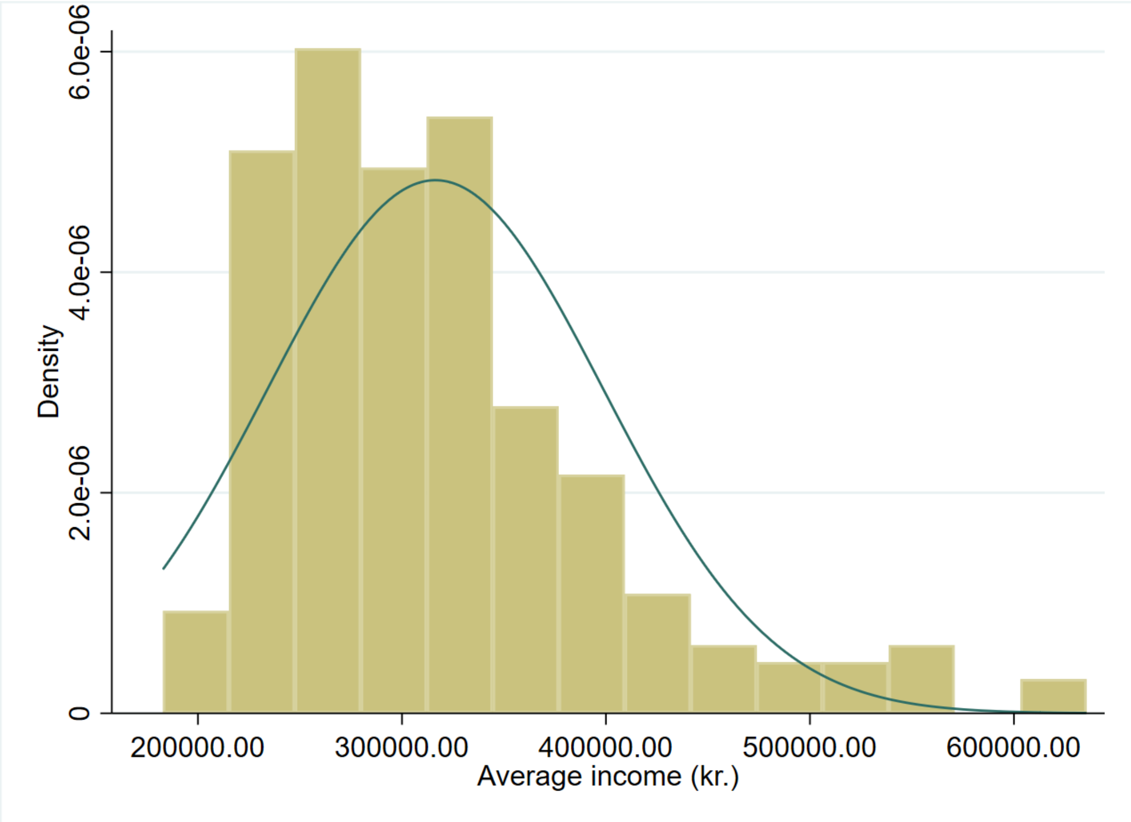
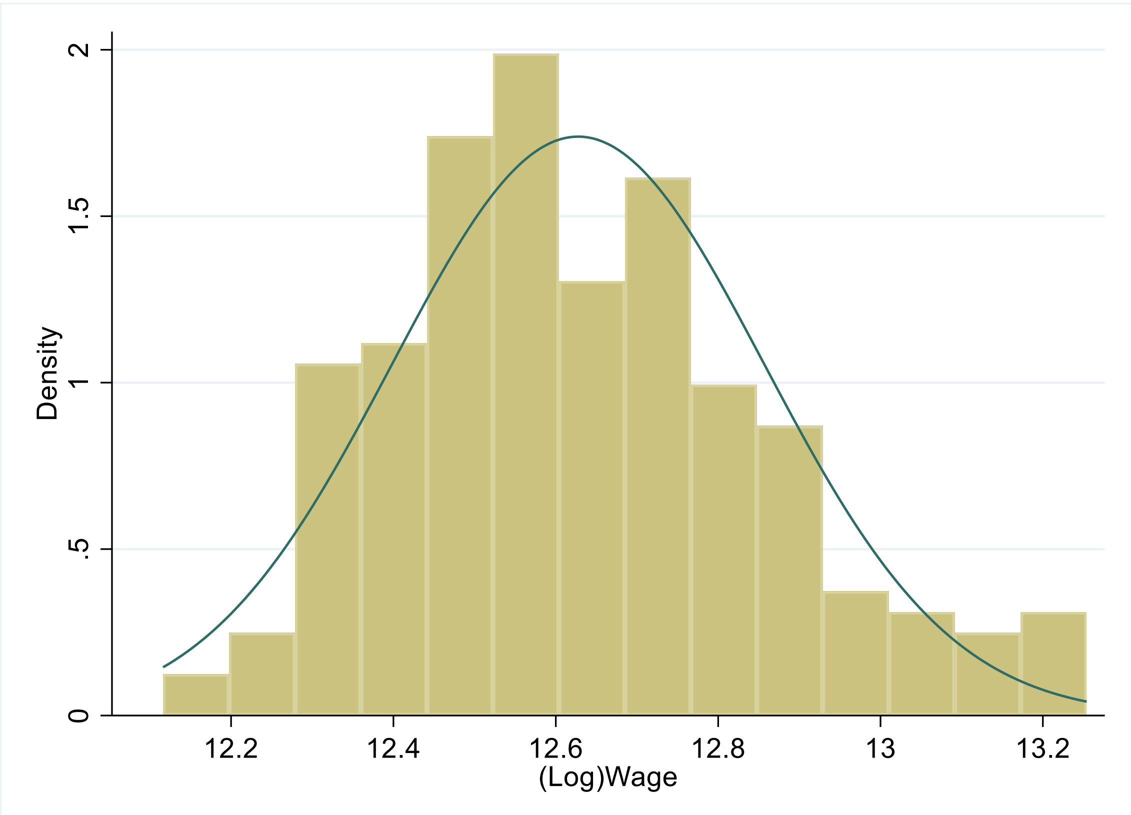


Figure 16. Service area at 300 meters for medium and large green spaces in Copenhagen – adapted from (The Technical and Environmental Administration, 2018).

Appendix B



A1. Distribution of income.



A2. Distribution of natural logarithmic transformation of income.

Appendix C

Table 13. Urban Atlas - land use classification and reclassification: Residential areas (red), green areas (green), other land use (grey).

Land use classifications: 2006	**Land use classifications: 2012 & 2018**
11100: Continuous urban fabric	11100: Continuous urban fabric
11210: Discontinuous dense urban fabric	11210: Discontinuous dense urban fabric
11220: Discontinuous medium density urban fabric	11220: Discontinuous medium density urban fabric
11230: Discontinuous low density urban fabric	11230: Discontinuous low density urban fabric
11240: Discontinuous very low density urban fabric	11240: Discontinuous very low density urban fabric
11300: Isolated structures	11300: Isolated structures
12100: Industrial, commercial, public, military and private units	12100: Industrial, commercial, public, military, and private units
12210: Fast transit roads and associated land	12210: Fast transit roads and associated land
12220: Other roads and associated land	12220: Other roads and associated land
12230: Railways and associated land	12230: Railways and associated land
12300: Port areas	12300: Port areas
12400: Airports	12400: Airports
13100: Mineral extraction and dump sites	13100: Mineral extraction and dump sites
13300: Construction sites	13300: Construction sites
13400: Land without current use	13400: Land without current use
14100: Green urban areas	14100: Green urban areas
14200: Sports and leisure facilities	14200: Sports and leisure facilities
20000: Agricultural areas	
	21000: Arable land (annual crops)
	22000: Permanent crops
	23000: Pastures
	24000: Complex and mixed cultivation patterns
	25000: Orchards
30000: Forests and semi-natural areas	
	31000: Forests
	32000: Herbaceous vegetation associations
	33000: Open spaces with little or no vegetations
	40000: Wetlands
50000: Water	50000: Water

Appendix D

Margins Methods and formulas. The following illustrates the **default prediction** for several of Stata's estimation commands (Stata, 2021). Margins computes estimates of:

$$p(\boldsymbol{\theta}) = \frac{1}{M_{S_\mathrm{p}}} \sum_{j=1}^{M} \delta_j(S_\mathrm{p}) f(\mathbf{z}_j, \boldsymbol{\theta})$$

Where), θ is the vector of parameters in the current model fit, **z** is a vector of covariate values, and $f(\mathrm{Z_j}\theta)$ is a scalar-valued function returning the value of the predictions of interest. $\delta_j(S_\mathrm{p})$ identifies elements within the subpopulation S_P (for the prediction of interest),

$$\delta_j(S_\mathrm{p}) = \begin{cases} 1, & j \in S_\mathrm{p} \\ 0, & j \notin S_\mathrm{p} \end{cases}$$

M_{S_P} is the subpopulation size, and *M* is the population size, as:

$$M_{S_\mathrm{p}} = \sum_{j=1}^{M} \delta_j(S_\mathrm{p})$$

Hence, let $\widehat{\theta}$ be the vector of parameter estimates, whereafter margins estimates $p(\theta)$via:

$$\widehat{p} = \frac{1}{w.} \sum_{j=1}^{N} \delta_j(S_\mathrm{p}) w_j f(\mathbf{z}_j, \widehat{\boldsymbol{\theta}})$$

where:

$$w. = \sum_{j=1}^{N} \delta_j(S_\mathrm{p}) w_j$$

$\delta_j(S_\mathrm{p})$ indicates whether observation j is in subpopulation S_p, w_j is the weight for the j^{th} observation, and *N* is the sample size.

For marginal effects of continuous covariate *x*, margins computes from the previous equation:

$$h(\mathbf{z}, \boldsymbol{\theta}) = \frac{\partial f(\mathbf{z}, \boldsymbol{\theta})}{\partial x}$$

The marginal effect for level *k* of factor variable *A* is the simple difference comparing its margin with the margin at the base level:

$$h(\mathbf{z}, \boldsymbol{\theta}) = f(\mathbf{z}, \boldsymbol{\theta} | A = k) - f(\mathbf{z}, \boldsymbol{\theta} | A = \mathrm{base})$$

Fixing covariates and balancing factors are done by controlling the values in each z vector through the marginlist, the at() option, among others. Suppose z is composed of the elements from the equation specification:

A##B x

where A is a factor variable with a levels, B is a factor variable with b levels, and x is a continuous covariate. To simplify the notation for this discussion, assume the levels of A and B start with 1 and are contiguous. Then:

$$z = (A_1, \ldots, A_a, B_1, \ldots, B_b, A_1B_1, A_1B_2, \ldots, A_aB_b, x, 1)$$

where A_i , B_j , A_iB_j represent indicator values for factor variables A and B and interaction A#B.

When factor A is in the *marginlist*, margins replaces A with *i* and then computes the mean of the subsequent prediction, for *i* = 1, . . . , a. When the interaction term A#B is in the *marginlist*, margins replaces A with *i* and B with j, and then computes the mean of the subsequent prediction, for all combinations of *i* = 1, . . . , a and *j* = 1, . . . , b.

The at() option sets model covariates to fixed values. Margins specified with at(x=value) causes margins to temporarily set x to that value for each observation in the dataset before computing any predictions. Thus, each z vector will look like

$$z = (1/a, \ldots, 1/a, B1, \ldots, Bb, B1/a, B2/a, \ldots, Bb/a, x, 1)$$

Standard errors are obtained by the delta method by default, which assumes that the values at which the covariates are evaluated to obtain the marginal responses are fixed. By default, margins uses the delta method to estimate the variance of $\hat{p}$:

$$\widehat{\mathrm{Var}}(\hat{p}|\mathrm{z}) = \mathbf{v}'\mathbf{V}\mathbf{v}$$

Here **v** is a variance estimate for $\widehat{\theta}$ as follows:

$$\mathbf{v} = \left.\frac{\partial\widehat{p}}{\partial\boldsymbol{\theta}}\right|_{\theta=\widehat{\theta}}$$

The variance estimate is conditional on the **z** vectors used to compute marginalized predictions.

Pwcompare calculates the margins as linear combinations of the coefficients. Let *k* be the number of levels for a factor term in our model. Then there are k margins for that term, and *m* unique pairwise comparisons of those margins.

$$m = \binom{k}{2} = \frac{k(k-1)}{2}$$

Appendix E

Table E1. Comparison of two-way linear interaction models – Fixed Effects

	Restricted	Simple	Theoretical
Net immigration	-.93454188	-4.1298136*	-.92975778
Green service Area	.2732018	.1335044	.13805714
Gini index	.03762378***	.03975131***	.01808422***
Net immigration # Green service Area	**-1.359796**	**2.8994903**	**-.56408975**
N 0-19 years %		1.0561677*	.69394474**
N 70-99+ years %		1.4760826**	1.0146187***
Social housing area / res		.00184893**	.00105687**
Higher education %			1.3665387***
Lower education %			-.76826195***
Unemployed %			-1.0768081**
Danish citizens %			.4916244***
Constant	11.357272***	11.037754***	11.338117***
R^2 within	.6454604	.69621871	.92258819
R^2 between	.52097179	.4463894	.78737173
R^2 overall	.55193264	.49598763	.81462671

*Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1*

Table E2. Comparison of three-way quadratic interaction models – Fixed Effects

	Simple	Parsimonious	Theoretical
Net immigration	-36.594613	-13.589796	9.0222521
Gini index	-.11531867	-.28276359	-.10075384
Net immigration # Gini index	54.694216	22.847628	-12.04677
Green service area	.03200771***	.03548161***	.01758586***
Net immigration # Green service area	1.3607586	.54834766	-.27146163
Gini index # Green service area	.01433391	.01269013	.00670553
Net immigration # Gini index # Green service area	-2.3292418	-1.0931657	.34176915
Immigration^2	722.96789	212.95307	-364.94248
Immigration^2 # Gini index	-1075.4455	-190.14519	716.21479
Immigration^2 # Green service area	-27.0346	-9.5447641	11.536161
Immigration^2 # Gini index # Green service area	**43.529147**	**11.641383**	**-23.81845**
N 0-19 years %		1.6287446***	.99869687***
N 70-99+ years %		1.173787**	.95330539***
Unemployed %		-2.2845801**	-1.2621525**
Social housing area / res		.00175315*	.00143079**
High education %			1.281011***
Lower education %			-.79460654***
Danish citizens %			.49843168**
Constant	11.520453***	11.153198***	11.330194***
R^2 within	.70320769	.76743063	.93392304
R^2 between	.51510745	.50612228	.799477
R^2 overall	.55540247	.54947148	.82427066

*Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1*

Appendix F

Table 14. Average marginal effects for simple slopes in the two-way quadratic interaction model (generalized for ten slopes).

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf. interval]	
Net immigration at slope:						
1	2.835	2.331	1.220	0.224	-1.734	7.403
2	1.743	1.867	0.930	0.350	-1.916	5.402
3	0.651	1.421	0.460	0.647	-2.135	3.437
4	-0.441	1.019	-0.430	0.665	-2.438	1.557
5	-1.533	0.735	-2.080	0.037	-2.974	-0.091
6	-2.625	0.725	-3.620	0.000	-4.046	-1.203
7	-3.716	0.997	-3.730	0.000	-5.671	-1.762
8	-4.808	1.395	-3.450	0.001	-7.543	-2.074
9	-5.900	1.839	-3.210	0.001	-9.504	-2.296

Simple slopes:					
1._at:	green = .05	4. at:	green = .35	7. at:	green = .65
2. at:	green = .15	5._at:	green = .45	8. at:	green = .75
3._at:	green = .25	6. at:	green = .55	9. at:	green = .85

Appendix G

Table 15. Green service areas at home-level. Adapted from (The Technical and Environmental Administration, 2018)

Service Area	Homes serviced		Homes underserviced	
	%	Number	%	Number
300 m	46,7 %	147.977	53,3 %	168.682
400 m	62,1 %	196.561	37,9 %	120.098
500 m	73,6 %	233.005	26,4 %	83.654
600 m	83,3 %	263.918	16,7 %	52.741
700 m	89,6 %	283.703	10,4 %	32.956
800 m	93,7 %	296.788	6,3 %	19.871
900 m	96,6 %	305.914	3,4 %	10.745
1000 m	98,2 %	310.979	1,8 %	5.680

Appendix H

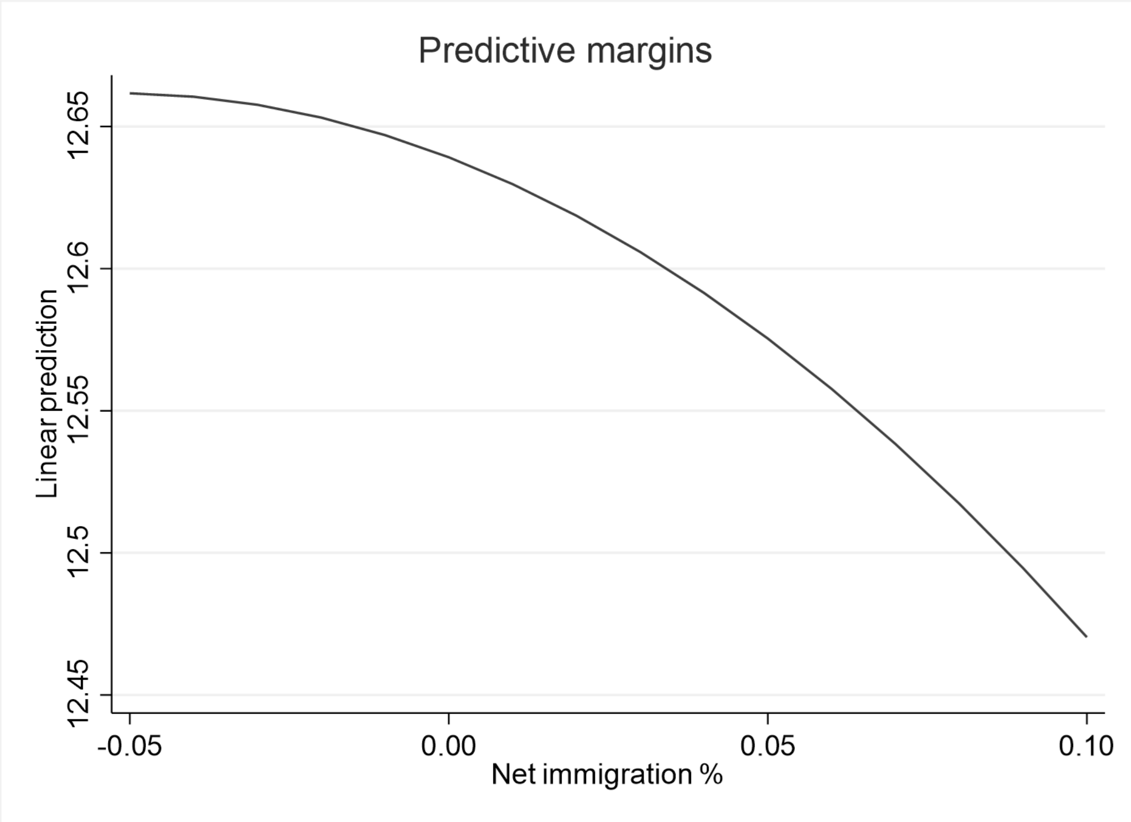


Figure 17. Predicted margins of net immigration - baseline without moderation.

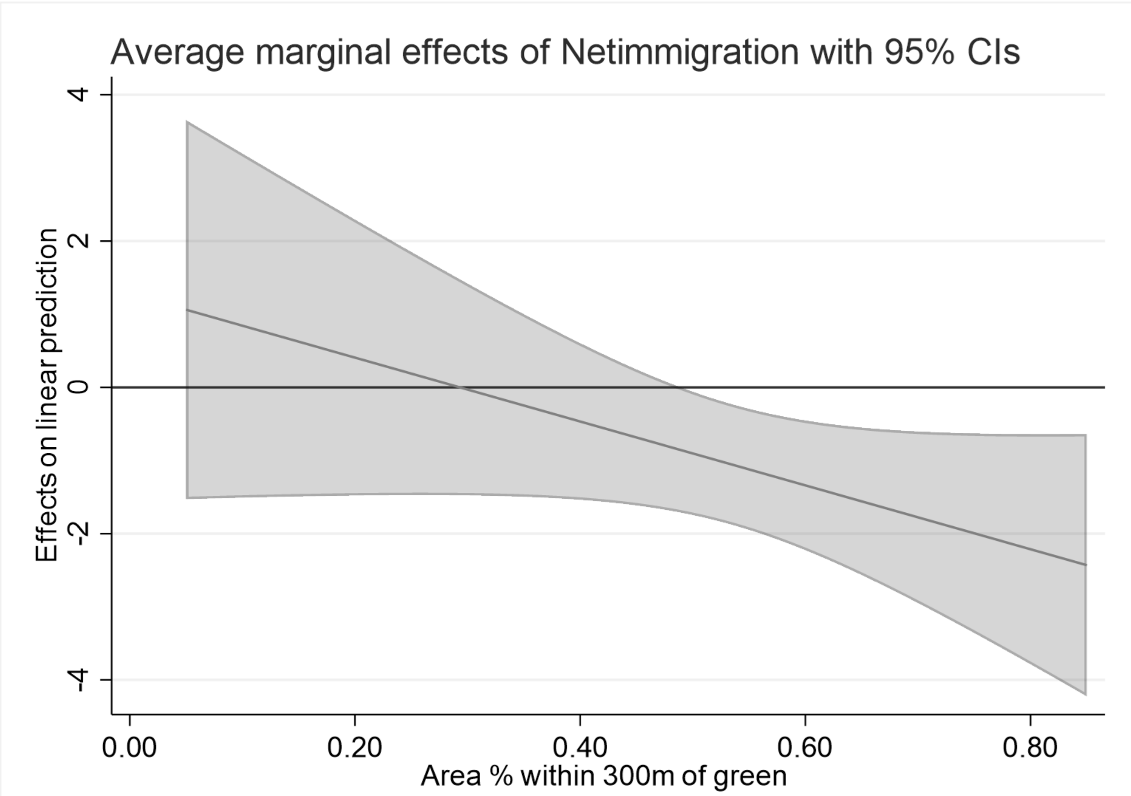


Figure 18. Average marginal effects of net immigration at all values of green service area.

Appendix I

Gender Ratio

The gender wage gap has long been an area of intensive investigation as gender differences in occupation and industry continue to be important in terms of wages (Kim & Polachek, 1994; Kunze, 2008). As such gender ratio is included as a control variable obtained from the dataset KKBEF1 and is normalized to obtain percentage variables for male and female residents.

Employment

The ratio of unemployed residents should be expected to have a notable impact on wage levels, as their incomes would be limited to social benefits, which is usually less than working salaries. The dataset KKARB1 is used to obtain the data and is normalized to obtain a percentage.

Ethnicity

The race wage gap is well known in research (Ananat, Shihe & Ross, 2018), and would be a natural addition to the control variables. However, the only available dataset containing all residents by ethnicity does not cover the full temporal scope of the study. An alternate variable is thus used, which combines the race wage gap with the well-researched citizenship wage gap (Steinhardt, 2012). The dataset KKBEF8 contains data on citizenship by ethnicity and should therefore be a good control variable. The dataset contains residents by Danish ethnicity, western ethnicity, and non-western ethnicity, which is normalized by number of residents.

Population Density

Studies widely indicate that densely populated cities enhance worker productivity via effects such as knowledge spillover and better labor market access (Masayuki, n.d.; Wheeler, 2004), and as such this variable is included as a potential control variable. Area size of neighborhoods is collected from KKAREAL, which is normalized by number of residents to obtain density.

Housing

Housing conditions are a crucial marker of socio-economic inequalities in Danish society (Damm, Schultz-Nielsen & Tranæs, 2016; Kristensen & Larsen, 2007). Therefore, KKBOL3 is used to create control variables for coverage of diverse housing conditions in the study area, including private ownership, rental, housing association, social housing, and public buildings. Of these options private ownership and social housing are taken as opposite ends of the relationship with wage, and these are used as control variables. Since the areas are not sharable and are subject to density, an area percentage of the total area is therefore not useful. Instead the areas are normalized by number of residents to get a more correct specification.

Table 16. Summary statistics of variables.

Variable	*Obs*	*Mean*	*Std. dev.*	*Min*	*Max*
Average income (kr.)	198	313229.4	76651.8	182816	570329
log wage	198	12.62756	0.22939	12.11624	13.25397
Net immigration %	198	0.01052	0.01299	-0.03503	0.093023
Green service area %	198	0.53055	0.159682	0.060777	0.841291
Gini index	198	30.37015	5.921243	20.48	48.62
Lower education %	198	0.28012	0.061625	0.131105	0.604651
Medium education %	198	0.28229	0.049672	0.176602	0.589147
Upper education %	198	0.203635	0.081444	0.018377	0.372081
N 019 years %	198	0.193878	0.037822	0.10828	0.325415
N 20-69 years %	198	0.732071	0.054314	0.559036	0.872612
N 70-99+ years %	198	0.074051	0.034983	0.002388	0.195268
Unemployed %	198	0.025192	0.010499	0	0.069767
Danish ethnicity %	198	0.854005	0.061916	0.506661	0.941282
Western ethnicity %	198	0.079772	0.041052	0.023401	0.373673
Non-western ethnicity %	198	0.066223	0.052981	0.014276	0.377351
Private housing area / res	198	17.44561	40.88234	0	553.1705
Private rental area / res	198	18.38004	38.49128	0	479.6124
Housing associations area / res	198	20.87345	28.1515	0	190.0219
Social housing area / res	198	13.61412	22.2866	0	149.1163
Public buildings area / res	198	1.029209	4.381854	0	42.96623
Population density	198	363.7327	1066.715	28.77628	11442.66
Area (km2)	198	1388654	1268560	146750	8265918
Residents	198	8350.389	5544.099	129	32088