

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis MSc Economics & Business

Specialisation Urban, Port, and Transport Economics

Unlocking the Green Potential: Analysing the Relationship Between Green Spaces and Labour Productivity in the Netherlands

Name student: Heleen Koch

Student ID number: 495169

Supervisor: prof.dr. F.G. van Oort

Second assessor: Jeroen van Haaren, MSc

Date final version: 30 June 2024

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

Abstract

This study examines the relationship between green space and labour productivity in the Netherlands using panel data from CBS, Statistics Netherlands. Results show that green space has no significant association with labour productivity in the Netherlands. However, significant associations have been found for specific types of green. A 1% increase agricultural surface is associated with a significant decrease of 0.19% in labour productivity. A 1% increase in recreational surface significantly correlates with an increase of 1.00% of labour productivity. No significant results were found for woodland and nature surface. Industry-specific analysis found that a 1% increase in green space significantly correlates with a 0.31% decrease in labour productivity of personal services, and no significant results for the business services and manufacturing industries. A 1% increase in agricultural surface is significantly associated with a decrease of 0.34% in labour productivity of the personal services industry, and no significant associations for the business services and manufacturing industry. A 1% increase in recreational surface significantly correlates with an increase in labour productivity of all three industries, with the largest positive association in the personal services industry: a 1.01%, followed by the manufacturing industry and lastly the business services industry. A 1% increase in woodland and nature surface only significantly correlates with an increase of 0.19% in labour productivity in the business services industry. Region analyses show no significant results for aggregate green. A 1% increase in agricultural surface relates to a 0.42% decrease in labour productivity outside of the Randstad; no significant results were found for the Randstad. A 1% increase in recreational surface relates to a 1.05% and 0.61% increase in labour productivity in the non-Randstad and Randstad area, respectively. A 1% increase in woodland and nature surface significantly correlates with a 0.35% decrease in labour productivity in the Randstad; no significant results were found for the non-Randstad area. As causality cannot be secured, it is recommended to interpret the results as association rather than true causal effect. For further research, it is recommended to analyse other potential externalities of changing green spaces. This will provide the government and other decisionmakers a more complete view on land development implications and reduce the risk of unforeseen consequences of policies.

Key words: green space, nature, labour productivity, industry analysis, regional analysis, panel data, the Netherlands, COROP.

Table of Contents

1. <i>Introduction</i>	6
2. <i>Literature Review</i>	8
2.1 The stress reduction theory and attention restoration theory	8
2.2 Effects of outdoor green on (mental) health.....	9
2.3 Productivity determinants and its relationship with (mental) health.....	11
2.4 Effects of outdoor green on labour productivity	12
2.5 Relevance and hypotheses.....	13
3. <i>Data</i>	16
3.1 Data collection.....	16
3.2 Descriptive statistics general analysis	19
3.3 Descriptive statistics industry-specific analyses	20
3.4 Descriptive statistics region-specific analyses	21
4. <i>Methodology</i>	23
4.1 Model estimation technique	23
4.2 General analysis	23
4.3 Industry-specific analyses	24
4.4 Region-specific analyses	25
4.5 Multicollinearity checks.....	25
5. <i>Results</i>	27
5.1 General analysis	27
5.2 Industry-specific analyses	29
5.2.1 Business services industry analysis.....	29
5.2.2 Personal services industry analysis	30
5.2.3 Manufacturing industry analysis	32
5.2.4 Comparison of the results of the different industry-specific analyses.....	34
5.3 Region-specific analyses	36
5.3.1 Randstad region analysis	36
5.3.2 Non-Randstad region analysis	38

5.3.3 Comparison of the results of the different region-specific analyses	40
6. <i>Discussion</i>	42
6.1 Hypotheses	42
6.2 Research limitations	44
6.3 Implications and recommendations	46
7. <i>Conclusion</i>	48
8. <i>Reference List</i>	49
<i>Appendix A Literature on the effect of indoor greenery on (mental) health</i>	58
<i>Appendix B Data description</i>	59
<i>Appendix C Detailed summary statistics</i>	61
<i>Appendix D Multicollinearity checks</i>	63

Overview of Tables

<i>Table 1 Overview of industry-specific models</i>	17
<i>Table 2 Overview of variables</i>	18
<i>Table 3 Descriptive statistics, general analysis</i>	19
<i>Table 4 Detailed descriptive statistics variable green space (in hectares)</i>	20
<i>Table 5 Descriptive statistics variable labour productivity in gross value added per FTE in euros, per industry-specific model</i>	21
<i>Table 6 Descriptive statistics per region-specific model</i>	21
<i>Table 7 Results of the general analysis with ln(labour productivity) as the dependent variable</i>	27
<i>Table 8 Results of the business services industry analysis with the natural logarithm of labour productivity of the business services industry as the dependent variable</i>	29
<i>Table 9 Results of the personal services industry analysis with the natural logarithm of labour productivity of the personal services industry as the dependent variable</i>	31
<i>Table 10 Results of the manufacturing industry analysis with the natural logarithm of labour productivity of the manufacturing industry as the dependent variable</i>	33
<i>Table 11 Comparison of the second models of the industry-specific analyses, with the natural logarithm of labour productivity of the industry mentioned in the model names</i>	34
<i>Table 12 Comparison of Model 3 of the three different industries analysed, with the dependent variable being labour productivity of the industry named in the model</i>	35
<i>Table 13 Results of the Randstad region analysis with the natural logarithm of labour productivity as the dependent variable</i>	37
<i>Table 14 Results of the non-Randstad region analysis with the natural logarithm of labour productivity as the dependent variable</i>	38
<i>Table 15 Comparison of the second models of the region-specific analyses, with the natural logarithm of labour productivity from the region as dependent variable</i>	40
<i>Table 16 Comparison of the third models of the region-specific analyses, with the natural logarithm of labour productivity as dependent variable</i>	41
<i>Table B1 Overview of codes and corresponding names of COROP (CR) areas, 2023</i>	59
<i>Table B2 Overview of soil types included in the variable green space</i>	59
<i>Table B3 Overview of educational level classifications</i>	60
<i>Table C1 Descriptive statistics per variable per model</i>	61
<i>Table D1-D12 Correlation matrices and Variance Inflation Factors</i>	63

1. Introduction

The Netherlands has a housing problem: there is a shortage of 390,000 houses (Mouissie & Kraniotis, 2023). Due to this shortage, people need to wait years to get suitable housing or cannot afford a house due to price increases. It has been determined that 981,000 houses need to be added to the current supply by 2030 (Ministerie van Algemene Zaken, 2023). The next question that comes to mind is where to build these homes. This debate has been going on for a while, as space is also needed for other purposes, such as nature or other buildings such as companies (NOS, 2021). Municipalities have not always been satisfied with decisions on housing development projects on their green areas (NOS, 2021).

Green areas have been proven to have multiple benefits. For example, green space has a positive impact on health (Maas et al., 2006), and nature enables attention restoration (Berto 2005). However, changes in current green areas could lead to externalities the Dutch government is currently unaware of. In an article of NRC, Aan de Brugh (2024) mentioned that quantifying and proving what the exact effect of greenery is, is very difficult. Research seems to identify a positive effect, but it is hard to estimate a true causal effect as many elements are relevant (Aan de Brugh, 2024). Furthermore, different types of greenery are often not analysed. Land development has been a difficult topic, as many people have different opinions, but what effects are proven to be true? This study aims to identify one of the potential externalities of changing green space, labour productivity, and analyses the relationship of types of green soils.

In the past, some research has been performed on the effect of greenery on the productivity of labour. Horr et al. (2016) found that nature in offices contributes to obtaining higher productivity. Furthermore, people are more satisfied with their working place if nature is accessible (Lottrup et al., 2013). However, most research focuses on the effects of green space on health and there remains a scarcity in quantifying the potential effects on productivity levels on a larger scale. As this study aims to close this gap, it is scientifically relevant. Furthermore, the results of this study could be beneficial for decision-making regarding land development, as this study tries to quantify the relationship between green space and a key performance indicator of the economy, labour productivity, which makes this study socially relevant as well. The results of this study can contribute towards optimising the location of companies, but also of houses. For example, if green surface has a positive association with labour productivity, the government might want to think twice before allocating green space to housing, as a decrease in green space could potentially result in a loss of labour productivity in that area. Therefore, the research question is as follows:

“What is the relationship between green surface and labour productivity in the Netherlands?”

To examine this research question, a quantitative analysis will be done with the use of panel data from the Netherlands between 1996 and 2017. The dependent variable is *labour productivity*, and the independent variables are (types of) *green space* and several control variables. For this research, *green space* has been defined as the total surface of recreational, agricultural, and woodland and nature; this definition is also used by Compendium voor de Leefomgeving (2023). *Labour productivity* is defined according to the definition set by the Central Bureau for Statistics (CBS): value added per volume (FTE) of labour (CBS, 2023b). Throughout this study, the results are interpreted as an association rather than a true causal effect as causality is not ensured. One of the limitations is that there might be a sorting effect – the more productive people work in the more green areas or the other way around – of which there could not be fully accounted for.

The research is on COROP level: a cluster of one or more adjacent municipalities in the Netherlands (CBS, n.d.). Most research has been performed on neighbourhood or city level; however, COROP level is the smallest level on which the data is available. With this regional division, the potential effect of green will be assessed on a new larger scale than previous literature. By using this division, it is also possible to account for differences in work- and living locations, as delimitation of COROP areas is based on commuter flows. Additionally, Raad voor de Leefomgeving en Infrastructuur et al. (2023) mention that welfare information often is available on COROP level, as these regions are more consistent compared to smaller regions, which often differ on a lot of measures between each other. Lastly, Raad voor de Leefomgeving en Infrastructuur et al. (2023) recommend that the government should consider the consequences of their actions for all regions, not just the core economic areas. Therefore, all COROP areas are considered for this research and a separate analysis will be done to look at differences in outcomes between the Randstad and non-Randstad area.

The next chapter, Chapter 2, provides an overview of existing literature. The data is described in Chapter 3, which includes the source of the data and the descriptive statistics. Chapter 4 explains the methodology to analyse the relationship and shows the regression equations. In Chapter 5, the results of the different analyses will be discussed. Chapter 6 includes a discussion on the hypotheses, limitations, implications, and recommendations. Lastly, the conclusion of this study is presented in Chapter 7.

2. Literature Review

Already a long time ago, the concept of *biophilia* was introduced by Wilson (1984). *Biophilia* suggests that human beings require interactions with nature because of evolutionary reasons, and that nature is an essential element of our existence (Wilson, 1984). Several research has been done on the effect of green space on health, ranging from the effects on stress (Fan et al., 2011; Luo & Jiang, 2022; Yang et al., 2020) to its positive effect on improving (mental) health (Braçe et al., 2020; Bringslimark et al., 2007; Lohr et al., 1996). An increased (mental) health of employees might result in more efficient workflows and less absenteeism in the workplace. This could improve the value added per employee: there might be a relationship between green spaces and labour productivity of companies. In this chapter, a comprehensive examination of existing research related to the chosen research topic will be discussed.

2.1 The stress reduction theory and attention restoration theory

There are multiple theories relevant for the research topic. The first one to be discussed is the *stress reduction theory*. According to European Parliament (2023), work-associated stress is a common cause for mental health issues in the European Union. In 2022, almost fifty percent of the respondents of a survey of EU-OSHA reported that they frequently faced extreme time constraints or had too much work to do. Mental health issues can be a cause for a decline in productivity and a rise in work absences (European Parliament, 2023). The *stress reduction theory* is about how mental stress and the subjection to environments are related (Ulrich, 1991). According to this theory, subjection to natural environments decreases mental pressure or stress, while on the contrary, exposure to urban surroundings can cause it. Ulrich (1991) states that this all relates back to the evolutionary past of humans in natural environments. Luo and Jiang (2022) applied this theory to a modern densely populated city: Hong Kong. The authors found that the feeling of being oppressed acts as an intermediary for urban-related views and mental stress. The main cause of someone feeling oppressed is urban-related views, for example skyscrapers, many vehicles, and other artificial appearances (Luo and Jiang, 2022). Furthermore, people associate dense cities often with chaos, which makes people overwhelmed, and giving a feeling of being out of control (Chung et al., 2019). As a result, people may experience stress (Luo & Jiang, 2022). Displeasure in densely populated and busy cities is a major issue. For example, Öhrström (2004) did research on the effect of traffic on people's wellbeing, and found that a decrease in road traffic, and therefore also a reduction in noise, has a positive effect on people's wellbeing.

Next to the *stress reduction theory*, there is also an *attention restoration theory* regarding nature. Kaplan's (1995) *attention restoration theory* states that subjection to natural environments serves as attention restoration and improves cognitive abilities. Natural settings are especially abundant in elements required for restoration (Kaplan, 1995). Contrary to natural surroundings, urban surroundings are full of distractions that require additional attention, and thus having an opposite effect to nature (Berman et al., 2008). Urban environments have a less recovering role in terms of attention than natural environments. In experimental research of Berto (2005), participants, a group of students, were only able to restore their attention after being mentally drained if they were exposed to restorative surroundings like nature, which was in the form of restorative pictures on a computer. Only people subjected to these natural environments enhanced cognitive abilities; the results of participants subjected to non-restorative environments were not significant and thus no evidence of attention restoration was found. Hartig et al. (2003) did research on the bridge between the *stress reduction theory* and the *attention restoration theory*. Compared to an urban environment, walking in a natural environment caused blood pressure to change in such way to allow for a larger stress reduction effect, which subsequently led to better scoring on an attention assessment (Hartig et al., 2003). The authors gathered the data with an experimental design with students. The walk took place at a natural park in a town next to their university city (Hartig et al., 2003). Berman et al. (2008) found the same results in their experimental study with students: students that walked in a park nearby the university scored better on various memorisation tests, such as a code which they had to remember in reverse order, compared to the group of students that had to take a walk in the city. Before the walk, they were mentally fatigued by an overload of information intake. The students in the group with nature walks showed superior ability in restoring their attention (Berman et al., 2008).

2.2 *Effects of outdoor green on (mental) health*

Several research has been done on the effects of outdoor green, ranging from the effect on levels of stress (Yang et al. 2020) to social cohesion (Cardinali et al., 2024). Many years ago, Ulrich (1984) already found evidence that hospital rooms with window views of nature shortened hospital stays. An experimental study was done in the US: half of the patients examined, was placed in a room with views of a brick building, and the other half in a room with views on nature. A primary finding of this research indicates that patients who had a room with views of nature, experienced a shorter postoperative hospital stay compared to the patients

with a brick wall view (Ulrich, 1984). Braçe et al. (2020) also found evidence on the effects of view on nature, but on mental health: a green space view from home can decrease feelings of depression and anxiety. A cross-sectional research was conducted with self-reported data from participants in a Spanish survey. The authors found that a green space view from home lowers the risk of anxiety and depression, but only for people which like the sight of green space from their residence (Braçe et al., 2020). They recommend increasing urban green spaces in residential areas to reduce the risks of anxiety and depression.

Next to the effects of having a view on green or nature, also several research has been conducted on the effects of being in or surrounded by green space. Taylor and Kuo (2009) found that children with ADHD are able to increase their attention with breaks consisting out of walks in a park already after twenty minutes, and that these types of breaks might improve concentration for other groups of the population as well. Fan et al. (2011) found that green surface in the neighbourhood lessens stress. The authors conclude that the surface of parks in the neighbourhood has a greater positive effect on (mental) health than for example backyards (Fan et al., 2011). The methodology used is a cross-sectional study with self-reported answers from a survey, which was conducted in Chicago, US. Next to this, Bratman et al. (2015) did an experimental study to compare the effect of walking in nature with the effect of an urban walk. Participants were randomly assigned to one of the groups, either to walk in nature or to do the urban walk. The results included that walking in nature causes a decrease in stress, and therefore maybe could restore positive affectivity. These results are in line with the stress reduction theory as well. Cardinali et al. (2024) even found that distinct features of green spaces, such as green space corridors within 800 metres, correlate with increased social cohesion, and indirectly lead to better mental health. Self-reported data on social cohesion and mental health of people living in four European cities was used. The authors conducted a sensitivity analysis including multiple indicators – such as level of greenness or vegetation, accessibility, and usage – on green surface features.

On larger scale than prior discussed area sizes, not much research has been done yet. Akpinar et al. (2016) did research on the effect of greenery, including different types of which it consists of, on mental health factors on ZIP code level within Washington State, USA. The ZIP code areas included in the dataset vary from 1.19 km² to 3,685.4 km² in size. No significant results were found for green space as a whole. However, inhabitants of ZIP code areas with relatively more woodland reported less days of mental health issues (Akpinar et al., 2016). Also, the size of woodland in the more densely populated areas such as cities was found to

decrease the number of days people reported issues concerning mental health. Richardson et al. (2011) also did research on a larger scale: city-level in the US. However, the authors did not find significant results between nature and mortality causes. The authors used a linear regression model to estimate the association (Richardson et al., 2011). In addition to outdoor greenery, several studies have analysed the potential effect of indoor greenery on (mental) health. However, as this study focuses on the relationship with outdoor green spaces, literature on the effect of indoor greenery can be found in Appendix A.

2.3 Productivity determinants and its relationship with (mental) health

Productivity is about the proportion of output compared to its input (Fried et al., 1993). Decreasing the time needed for an employee to produce an amount of output is the bottom line of productivity improvements (Fedulova et al., 2019). Levels of productivity can vary due to various reasons, e.g. technology differences, processes, and operational environments (Fried et al., 1993). Anos-Casero and Udomsaph (2009) state that a company's environment, including labour quality, has a significant effect on the productivity levels of a company. One of the factors determining labour productivity is human capital (Korkmaz & Korkmaz, 2017). Deterioration of health could for example decrease the size of the potential labour market, and thereby decrease the labour pool (Chansarn, 2010). One of the most important determinants of human capital is education (Huffman, 2001). Education results in obtaining knowledge advancements, which can alter productivity levels (Weisbrod, 1962). Adedeji and Campbell (2013) state that human capital is strongly dependent on the level of education, where a higher level increases human capital more. Illness and bad health are another key element impacting people's capabilities and actions, and employee productivity levels are directly impacted by employees' state of health (Bloom & Canning, 2003). For example, the World Health Organization (2001) uses employee productivity to advocate for higher health budgets for developing nations. Next to human capital, fixed capital also has a positive relationship with labour productivity (Trpeski et al., 2019)). Labour productivity increases are for a large part due to capital investments (Mankiw et al., 1992). Furthermore, densely populated areas, such as metropolitan areas, greatly increase employees' productivity levels, as knowledge can be easier shared, and a larger labour pool is available (Puga, 2010). A higher labour volume could indicate a more densely populated area.

2.4 Effects of outdoor green on labour productivity

Already some evidence has been found on the relationship between outdoor green and certain abilities which as result might increase labour productivity as well. In a study regarding nature and cognitive ability, an association was found between distance from home to nature, including parks and recreational surface, and time to finish a test (Zijlema et al., 2017). Data of multiple regions in Europe, including the Netherlands, was used for this cross-sectional study. Per 100-meter increase in the distance, the time to complete a neuropsychological exam increased with 1.50 percent; someone's cognitive abilities might profit from the nearness of natural areas. The cause of this relationship could not be determined due to insignificance of the possible cause relationships. The research of Bratman et al. (2015) found that a walk in a park compared to a walk in an urban area with a lot of traffic increases one's ability to remember spoken information. Furthermore, Tennessen and Cimprich (1995) found that students having natural environments as window views performed better during attention related exams than students with less natural views; this is also in line with the *attention restoration theory*. This study shows that views of a natural environment improve attention levels. Research has also been done on the effects of other climate or nature factors on productivity levels. With the use of an ordinary least squares technique, Yildirim et al. (2009) found that an increase in temperature leads to lower labour productivity.

Regarding labour productivity among different industries, it is important to look at the capital-labour ratio. One of the two determinants of labour productivity is the capital-labour ratio, which measures capital intensity (Guest, 2011). Overall, the capital-intensive industries have higher value added per employee as more capital is used compared to labour-intensive industries (Guest, 2011). As for example cognitive ability increases with (exposure to) green space (Berto, 2005; Kaplan, 1995; Zijlema et al., 2017), a change in green space might cause the largest change in labour productivity of labour-intensive industries. In the article of Guest (2011), it is mentioned that personal services, such as personal care, are relatively more labour intensive. According to Dieppe & Matsuoka (2021), the agricultural industry contributes to low labour productivity in emerging markets and developing economies. People have been switching from agricultural industries to other more productive sectors, causing a rise in productivity of countries (Dieppe & Matsuoka, 2011).

2.5 Relevance and hypotheses

Russell et al. (2013) state that much research has been performed on the relationship between green on health outcomes. However, the effect of green on other outcomes – such as learning, knowledge, and motivation – has mainly been researched theoretically and there remains a scarcity in quantifying the relationship (Russell et al., 2013). Some of the studies are based on self-reported data, such as surveys, which could decrease the reliability if for example respondents interpret the questions differently. Furthermore, a relationship between green surface and labour productivity in value added per unit of labour has not been assessed yet. It could be interesting for companies, but also governmental bodies to know whether there is an association between green and the value added per unit of labour. It contributes to decision-making of not only location optimisation of companies and green spaces, but to policy making as well. Therefore, in this thesis the relationship between (different types of) green and labour productivity will be assessed. The relationship will also be assessed on industry-level to see whether it is different across industries, as there are variations in work settings (e.g. working indoors vs. outdoors and working at the client vs. at the company office). Additionally, the relationship will be assessed on two regional samples of the dataset, Randstad and non-Randstad area, to assess whether the relationship is different across these two regions in the Netherlands. Based on the literature review, the following four hypotheses are formulated:

Hypothesis 1: An increase in green surface is associated with an increase in labour productivity in the Netherlands.

According to many studies, (types of) green spaces have a positive effect on (mental) health (Braçe et al., 2020; Cardinali et al., 2024; Fan et al., 2011; Ulrich, 1984). Labour productivity enhancements are dependent on the average time needed per output for a worker (Fedulova et al., 2019). According to Korkmaz & Korkmaz (2017), human capital is one of the main determinants of labour productivity. Bloom & Canning (2003) state that someone's abilities are impacted by one's health state, and therefore health has a direct effect on labour productivity. By combining the findings of existing literature, a positive relationship between green space on labour productivity is expected.

Hypothesis 2: Looking at the three different types of soil which together form green space: agricultural surface is expected to have a negative association with labour productivity, while

recreational, and woodland and nature surface are expected to have a positive association with labour productivity in the Netherlands.

In green space, agricultural land is included. Large agricultural surface in a COROP area could imply that a large part of the COROP area is being used for the agricultural industry, an industry with a relatively quite low labour productivity, see Table C1. Also, according to Dieppe & Matsuoka (2021), the agricultural industry could contribute towards low productivity levels. Furthermore, according to Compendium voor de Leefomgeving (2023), agricultural surface has been decreasing and the other types of green soils have been increasing, while labour productivity has been increasing in the Netherlands between 1996 and 2017 (CBS, 2023a). Therefore, a third model is included in the research which analyses the three different types of green soils separately which together form the variable green space: agricultural, recreational, and woodland and nature surfaces. By isolating agricultural surface, it is expected that an increase in the other two types of green space, a) woodland and nature and b) recreational surface, correlates with higher labour productivity. A study in the US found that woodland contributes to better mental health, and that size of the woodland area matters as well (Akpınar et al., 2016). Furthermore, according to multiple studies, having access to or walking in parks has a positive effect on health and contributes towards attention restoration (Berman et al., 2008; Berto, 2005; Hartig et al., 2003; Taylor and Kuo, 2009). Public green spaces, such as parks, are included in recreational surface, see Table B2. An increase in green types of soil in a certain COROP could imply that the chance of being exposed to it increases.

Hypothesis 3: The association of green surface with labour productivity differs across industries, where the largest positive association is expected in the most labour-intensive industry, the personal services industry, followed by the business services industry, and the smallest positive association is expected in the most capital-intensive industry: manufacturing.

As Guest (2011) mentioned that capital-intensive industries tend to have higher productivity levels due to their capital investments, it is expected that when a change in green space is observed, it may have the largest effect on the more labour-intensive industries, as the productivity levels of those industries are more sensitive to changes in human capital. As green space is expected to impact capabilities and mental well-being of employees, it is expected to change human capital and thus to have a larger positive correlation in the labour-intensive

industries. Understanding of the distinctions of how green correlates with labour productivity across industry types could bring insights that are essential to location strategies of companies and public bodies. If the personal services industry has a larger positive relationship with green space compared to the more capital-intensive industry, the government might want to introduce a policy to locate companies within the personal services industry in the greener areas and the more capital-intensive companies in the less green areas if the aim is to reach higher productivity levels in the Netherlands.

Hypothesis 4: A larger positive association between green and labour productivity is expected in the Randstad compared to the non-Randstad area in the Netherlands.

Many studies found positive relationships between nature and someone's health and abilities. According to Luo and Jiang (2022) people feel more under pressure with urban-related views. The attention restoration theory argues that nature improves the mind's ability in restoring attention (Berman et al., 2008; Kaplan, 1995), and walking in a natural environment decreases stress (Hartig et al., 2003). As green space is scarcer in the Randstad – the more urban area of the two as the four largest cities of the Netherlands are located in the Randstad – an increase in green space is expected to have a larger positive association with labour productivity in the urban area the Randstad compared to the non-Randstad more rural area. If it turns out that the relationship is larger in the Netherlands, the government knows where differences in green space has the highest association with labour productivity, and land development might be better in the areas where green spaces are less associated with labour productivity, to minimise potential productivity losses.

3. Data

For this research, an empirical research method is used in combination with secondary data, as an analysis is performed on existing data. The statistical software Stata/MP 17 is used to perform the analyses. This chapter discusses the data collection and descriptive statistics.

3.1 Data collection

The data for this research is from CBS. All data used is on COROP level. This scale level differs from the literature review, as this is the smallest area on which productivity data is available in the Netherlands and to be an extension of existing literature on smaller scales. COROP-level data will be used to account as much as possible for the difference in work and living areas. COROP areas are originally based on the commuter flows (Rijksinstituut voor Volksgezondheid en Milieu, n.d.). It could be that not only green surface at the work location contributes towards labour productivity, but also green at someone's home location. By using COROP-level data, we assume that work and living is in the same COROP. An overview of the forty COROP areas, including the corresponding numbers and names can be found in Table B1.

Four different datasets from CBS are used to create one dataset; these datasets are (1) soil use (CBS, 2023a), (2) production process (CBS, 2023d), (3) investments in fixed assets (CBS, 2023c), and (4) working population (CBS, 2014). The data for this research is panel data, as the dataset includes data on different regions in the Netherlands over time. The CBS database on soil use contains information on green space in hectares per type of soil on COROP level for the years 1996, 2000, 2003, 2006, 2008, 2010, 2012, 2015, and 2017 (CBS, 2023a). Three types of soil from this database are summed to get the surface of green space in hectares as defined: recreational surface, agricultural terrain, and woodland and natural terrain. A further clarification of what types of soil belong to these three types can be found in Table B2. The dataset on production process contains data on gross value added and total labour volume in labour years (FTE) on industry-level and per COROP between 1995 and 2021 (CBS, 2023d). Gross value added is divided by the total volume of labour in labour years (FTEs) to calculate labour productivity. Data of a specific COROP is based on the companies located within that COROP. The third dataset contains data on the investments in fixed assets in millions of euros on industry-level and per COROP between 1995 and 2021 (CBS, 2023c). The last dataset contains data on the educational level of the employed labour force, which includes people between the age of 15 and 65 years that have paid work and are working 12 hours or more per

week (CBS, 2014). The educational levels are classified into three levels: low, middle, and high. A detailed overview of the definitions of the different levels can be found in Table B3. Educational levels of the working population are available for the years between 1996 and 2013 on COROP level, but not on industry-level (CBS, 2014). For this study, the educational levels will be expressed in percentages of the total number of the employed labour force (i.e. the sum of the number of people in the three educational levels). Only the percentage of the working population with high obtained education will be used in this research; a decrease in this percentage implies that the percentage of the working population with low and middle obtained education increases.

If a certain year was not available, this is stated in the descriptive statistics section. The variables of the datasets are merged to one dataset to conduct the regression analyses. There will be one analysis performed to analyse the relationship without industry and region specifications, referred to as the *general model*, and thereafter multiple industry-specific and region-specific analyses will be performed to see whether the association is different across industries and regions. Table 1 shows an overview of the industries which will be analysed separately. The different regions that will be analysed are the Randstad and the non-Randstad area, Table B1 shows which COROP areas are part of the Randstad and which are not, according to the division set by CBS (2022).

Table 1 Overview of industry-specific models

Model name	Industries included (SIC 2008 industry code)
Business services	Information and communication (J); financial institutions (K); renting, buying, and selling of real estate (L); other specialised business services (M); renting and other business support (N).
Personal services	Wholesale and retail trade (G); transportation and storage (H); accommodation and food serving (I); public administration and services (O); education (P); health and social work activities (Q); culture, sports, and recreation (R); other service activities (S); households (T); extraterritorial organisations (U).
Manufacturing	Manufacturing (C)

Table 1 shows the industry-specific models and a brief description of what is included. Each of these models will be individually analysed to look at the relationship between green surface and labour productivity, and to see whether a difference in association between industries is observed. Industries are according to the Standard industrial Classifications (SIC) of 2008 the European Union.

Source: CBS (2023d).

For each of the models, i.e. the general, industry-specific, and region-specific models, the same variables are available. Table 2 shows an overview of the variables.

Table 2 Overview of variables

Variable name	Description
Region code	Number of COROP
Region name	Written out name of COROP, e.g. “Oost-Groningen (CR)”
Year	Year of observation
Id	Unique identification number of observation, combination of year and region code, e.g. for CR01 and year 1996 this is CR011996
Green surface	Surface of green space in hectares
Agricultural surface	Agricultural surface in hectares
Recreational surface	Recreational surface in hectares
Woodland and nature surface	Woodland and nature surface in hectares
Gross value added	Gross value added in millions of euros
Investments	Investments in fixed assets in millions of euros
Labour productivity	Gross value added per FTE in euros
Labour volume	Number of FTEs in thousands
High education	Percentage of employed people with high obtained education

Table 2 shows a concise overview of the variables included in the merged dataset. For industry level specific data of a certain variable, a new variable specific for this industry is created, i.e. for investments in the business services industry this variable is called Investments_business.

The variables that are used as control variables are total labour volume (CBS, 2023d), three different levels of education (CBS, 2014), and investments in fixed assets (CBS, 2023c). Total labour volume (in thousands of FTEs) is used to account for potential benefits of high dense areas, as higher labour volume could indicate that there are economies of scale, knowledge spill overs, higher availability of resources, and a larger labour pool (Puga, 2010). Percentage of working people with high obtained education is used to account for regional differences in educational levels which could also lead to differences in labour productivity (Adedeji & Campbell, 2013; Huffman, 2001; Weisbrod, 1962). According to CBS (n.d.-b), most people with high obtained education live in areas in which universities are located and in the north of the Randstad. Lastly, the variable investments in fixed assets (in millions of euros) is used as control variable as this variable is expected to account for the possible value added due to capital investments, so that the relationship of green space purely captures the input of human capital.

3.2 Descriptive statistics general analysis

The dataset includes thirteen variables which are relevant for the general analysis, which are the variables shown in Table 2. The dataset contains out of forty COROP areas and nine different years. The minimum and maximum number of waves in the dataset are equal, namely nine. The panel data is strongly balanced, i.e. for 100% of the units all waves are observable. Table 3 shows the summary statistics for the dependent and independent variables of the general model.

Table 3 Descriptive statistics, general analysis

Variable	Obs.	Mean	Std. dev.	Minimum	Maximum
Green surface	360	71,681.56	42,880.96	7,735.00	164,334.00
Agricultural surface	360	57,043.06	33,770.74	2,122.00	125,381.00
Recreational surface	360	2,438.48	1,529.19	337.00	7,853.00
Woodland and nature surface	360	12,200.01	13,353.67	359.00	78,206.00
Gross value added	360	13,039.23	14,590.70	1,056.00	100,385.00
Investments	360	3,111.00	3,676.53	225.00	39,052.00
Labour productivity	360	70,796.95	15,260.13	37,396.12	129,923.95
Labour volume	360	171.80	156.45	14.70	851.00
High education	280	0.28	0.08	0.11	0.50

Table 3 shows the summary statistics for the general model. As educational levels are not available for the years 2015 and 2017, the number of observations of high education is lower compared to the other variables.

All variables are continuous variables and have 360 observations, except the variable on high obtained education. The average of green surface of the dataset is 71,681.56 hectares. The minimum surface of green is 7,735.00 hectares and the maximum value is 164,334.00. The mean of agricultural surface is 57,043.06 hectares, with a minimum of 2,122.00 and a maximum of 125,381.00 hectares. Recreational surface has a mean of 2,438.48, a minimum of 337.00 and a maximum of 7,852.00 hectares. Woodland and nature surface has a mean of 12,200.01 hectares, a minimum of 359.00 and a maximum of 78,206.00 hectares. Gross value added has a mean of 13,039.23 in millions of euros, with a minimum of 1,056.00 and a maximum of 100,385.00 in millions of euros. The investments in fixed assets are on average 3,111.00 million euros, with a minimum value of 225.000 and a maximum of 39,052.00 in millions of euros. The average of labour productivity (gross value added per FTE) is 70,796.95 euros. The minimum is 37,396.12 and the maximum is 129,923.95 euros. Labour volume has a mean of 171.80 in thousands of FTEs, with a minimum of 14.70 and a maximum of 851.00

FTEs in thousands. The mean of the percentage of the working population with high obtained education is 0.28, with a minimum of 0.11 and a maximum of 0.50.

Table 4 shows the detailed panel data summary statistics for the (types) of green soils. It is important for these variables to change over time, as otherwise it will be difficult to identify potential associations with labour productivity. According to this table there is within variance, implying that within COROP areas the values of these variables change over time.

Table 4 Detailed descriptive statistics variable green space (in hectares)

Surface variables	Variation	Mean	Std. dev.	Minimum	Maximum	Obs.
Green	Overall	71,681.56	42,880.96	7,735.00	164,334.00	N = 360
	Between		43,282.94	7,872.22	161,976.90	n = 40
	Within		2,666.64	59,975.56	95,318.89	T = 9
Agricultural	Overall	57,043.06	33,770.74	2,122.00	125,381.00	N = 360
	Between		34,059.78	2,201.44	120,722.30	n = 40
	Within		2,499.02	47,349.62	76,538.51	T = 9
Recreational	Overall	2,438.48	1,529.19	337.00	7,853.00	N = 360
	Between		1530.43	355.89	7,107.11	n = 40
	Within		220.00	1,578.59	3,184.37	T = 9
Woodland and nature	Overall	12,200.01	13,353.67	359.00	78,206.00	N = 360
	Between		13,493.22	462.67	77,833.44	n = 40
	Within		557.34	8,055.90	16,223.90	T = 9

Table 4 shows the detailed summary statistics for the (types of) green space. These variables are the same for the industry-specific analyses, as these surfaces are only dependent on the COROP area.

By using the tabulate command in Stata the frequencies of values of the variables were checked. For the variable green surface, it was found that almost every value has a frequency of one, only two values have a frequency of two. All values of agricultural land have a frequency of one, thirteen values of recreational land have a frequency of two and all others values a frequency of one, four values of woodland and nature have a frequency of two and the other values a frequency of one. This implies that within a COROP area none of the green soil variables stay the same over the total time span analysed.

3.3 Descriptive statistics industry-specific analyses

Table 5 shows the descriptive statistics of the dependent variable labour productivity per industry. A detailed overview of all models' descriptive statistics can be found in Table C1.

Table 5 Descriptive statistics variable labour productivity in gross value added per FTE in euros, per industry-specific model

Labour productivity	Observations	Mean	Standard deviation	Minimum	Maximum
Business services	360	87,659.94	17,867.95	38,125.00	158,674.56
Personal services	360	56,769.24	11,321.33	32,370.13	88,128.55
Manufacturing	360	87,110.45	25,376.57	33,942.86	165,460.98

The mean of labour productivity of the general model is 70,796.95 euros per FTE. As can be seen in Table 5, the business services industry has the highest mean of 87,659.94 euros per FTE compared to other industries but is close to the mean of the manufacturing industry. The personal services industry has the lowest mean of 56,769.24 euros per FTE. The range of labour productivity values is the largest within the manufacturing industry, and the smallest within the personal services industry.

3.4 Descriptive statistics region-specific analyses

Table 6 shows the descriptive statistics of the region-specific analyses. The dataset is split into two subsamples: the Randstad data only includes COROPs located in that area, and the non-Randstad only includes data of COROPs that are outside of the Randstad.

Table 6 Descriptive statistics per region-specific model

Variable	Obs.	Mean	Std. dev.	Minimum	Maximum
Green surface					
Randstad	99	35,683.89	32,279.02	7,735.00	109,947.00
Non-Randstad	261	85,335.85	38,283.51	9,511.00	164,334.00
Agricultural surface					
Randstad	99	27,882.78	26,144.95	2,122.00	84,453.00
Non-Randstad	261	68,103.86	29,503.87	4,050.00	125,381.00
Recreational surface					
Randstad	99	2,677.12	2,188.72	508.00	7,853.00
Non-Randstad	261	2,347.96	1,180.37	337.00	5,434.00
Woodland and nature surface					
Randstad	99	5,123.99	5,643.20	359.00	21,161.00
Non-Randstad	261	14,884.02	14,419.40	542.00	78,206.00
Gross value added					
Randstad	99	22,795.46	22,861.80	2,311.00	100,385.00
Non-Randstad	261	9,338.58	6,852.46	1,056.00	34,267.00

Investments					
Randstad	99	5,321.62	6,025.31	646.00	39,052.00
Non-Randstad	261	2,272.49	1,553.42	225.00	7,088.00
Labour productivity					
Randstad	99	77,275.63	14,525.78	46,533.33	117,961.22
Non-Randstad	261	68,339.52	14,831.30	37,396.12	129,923.95
Labour volume					
Randstad	99	272.92	235.82	49.20	851.00
Non-Randstad	261	133.45	86.40	14.70	369.40
High education					
Randstad	77	0.34	0.08	0.18	0.50
Non-Randstad	203	0.26	0.06	0.11	0.45

Table 6 shows the summary statistics for the region-specific analyses. As educational levels are not available for the years 2015 and 2017, the number of observations is lower compared to the other variables.

As can be seen in Table 6, the mean of green surface is lower in the Randstad, implying that the average number of hectares of green space across all years and COROPs is lower in the Randstad compared to the non-Randstad area. The same holds for agricultural surface and woodland and nature surface. Only recreational surface has a higher mean in the Randstad; however, the mean of recreational surface does not differ that much between the Randstad and the non-Randstad area as the other types of green space. Gross value added has the highest average in the Randstad and the largest range of values. Investments in fixed assets also has the highest mean in the Randstad, and a much larger range is observed in the Randstad compared to the non-Randstad area. Labour productivity also has the highest mean in the Randstad; however, the maximum value is the highest in the non-Randstad area, and the range is the largest in this non-Randstad area. Labour volume is also the highest in the Randstad, with a larger range in values as well in that area. Lastly, looking at the percentage of employed with high obtained education, the mean is higher in the Randstad: the mean of the working population with high obtained education is 34% in the Randstad, compared to 26% in the non-Randstad area. This implies that on average the obtained educational level is higher in the Randstad. Furthermore, the minimum value is lowest in the non-Randstad area with 0.11, and the maximum value is highest in the Randstad with 0.5, implying that 50% of the working population has high obtained education in one of the observations of the Randstad.

4. Methodology

The aim of this research is to quantify the relationship between green space and labour productivity in the Netherlands. As for this research panel data is used, panel data estimation techniques are used to estimate the association. This chapter discusses the used estimation techniques and the equations for the different analyses.

4.1 Model estimation technique

The use of panel data makes it possible to better control for unobserved individual heterogeneity compared to a time series or cross-sectional study (Hsiao, 2003). For panel data, often two models are considered: the Fixed Effects and Random Effects model (Menard, 2008). The choice between panel data estimators is based on unbiasedness and efficiency. For the Pooled OLS estimation technique, one of the assumptions includes that there is no serial correlation (Menard, 2008). This assumption is highly unlikely to hold with panel data, as errors are serially correlated over time; for each observation of an individual, the effect of unobserved unit heterogeneity is similar (Menard, 2008). As this makes Pooled OLS inefficient and the standard errors incorrect, the Pooled OLS model will not be considered for this study.

To choose between a Fixed Effects and a Random Effects model, the Hausman test will be performed, which tests whether the differences in time-variant coefficients are significant or not (Amini et al., 2012). If the test is not significant and thus the null hypothesis cannot be rejected, a Random Effects model will be more efficient. In that case, the time-invariant characteristics do not cause a significant difference between the two models and therefore the model does not have to account for that. If the test is significant, the null hypothesis should be rejected and a Fixed Effects model will be used to eliminate the time-invariant characteristics, as the Hausman test showed that the time-invariant characteristics do matter.

4.2 General analysis

The dependent variable is labour productivity. The independent variable is surface of green space of the COROP in which the company is located. For the dependent variable labour productivity and independent variable green space, the natural logarithm is taken, as we are interested in the percentage change of labour productivity if green space changes with one percent. Equation 1 shows the model equation:

$$\ln(P_{it}) = \beta_0 + \beta_1(\ln(G_{it})) + \beta_2L_{it} + \beta_3E_{high,it} + \beta_4I_{it} + \alpha_i + \varepsilon_{it} \quad (1)$$

where P_{it} and G_{it} are the productivity of labour and surface of green of COROP i in year t , respectively. L_{it} , $E_{high,it}$, and I_{it} are the control variables labour volume, percentage of people in the working population with high obtained education, and investments in fixed assets of COROP i in year t , respectively. The COROP fixed effects are captured with α_i and ε_{it} captures the idiosyncratic error, the time-varying component of the error term. In case the Hausman test is significant, and a Fixed Effects model will be used, the COROP fixed effects part of the equation will be eliminated by time demeaning the data, as the unobserved fixed effects do not change over time. In addition to Equation 1, in Equation 2 the different types of soil of which green space consists of will be added to look whether the relationship is different across types of (green) soil.

$$\begin{aligned} \ln(P_{it}) = & \beta_0 + \beta_1(\ln(A_{it})) + \beta_2(\ln(R_{it})) + \beta_3(\ln(W_{it})) \\ & + \beta_4L_{it} + \beta_5E_{high,it} + \beta_6I_{it} + \alpha_i + \varepsilon_{it} \end{aligned} \quad (2)$$

where G_{it} has been split up into A_{it} , R_{it} , and W_{it} . These variables capture agricultural surface, recreational surface, and woodland and nature surface, respectively. The other variables remain the same as in Equation 1.

4.3 Industry-specific analyses

Next to the general analysis, multiple regressions will be performed per industry – business services, personal services, and manufacturing – to see whether the coefficient significantly differs across industries. For each industry identified, as in Table 1, an analysis will be performed. Equation 3 shows the regression for the labour productivity levels on industry level. As green surface is measured per COROP it does not differ across industries within that COROP.

$$\ln(P_{ijt}) = \beta_0 + \beta_1(\ln(G_{it})) + \beta_2L_{ijt} + \beta_3E_{high,it} + \beta_4I_{ijt} + \alpha_{ij} + \varepsilon_{ijt} \quad (3)$$

where P_{ijt} and G_{it} are the productivity of labour and surface of green of COROP i , industry j (b for business services, p for personal services, and m for manufacturing) in year t ,

respectively. L_{ijt} and I_{ijt} are the control variables labour volume and investments in fixed assets of COROP i , industry j in year t , respectively. $E_{high,it}$ is the control variable with the percentage of people in the working population with high obtained education of COROP i in year t . As educational levels are not available on industry-level, this control variable is the same as in the general model as it is only dependent on COROP region. α_{ij} are the COROP area fixed effects for industry j and ε_{ijt} captures the idiosyncratic error, the time-varying component of the error term. As well in these analyses, in case the Hausman test is significant and thus a Fixed Effects model will be used, the COROP area fixed effects (α_{ij}) will be eliminated by time demeaning the data. As well for the industry-specific analyses another regression will be performed on the three different types of soil, to see whether a different relationship with labour productivity is observed across types of green soil, see Equation 4. The different types of soil are not dependent on industries, only on the COROP area itself.

$$\ln(P_{ijt}) = \beta_0 + \beta_1(\ln(A_{it})) + \beta_2(\ln(R_{it})) + \beta_3(\ln(W_{it})) + \beta_4 L_{ijt} + \beta_5 E_{high,it} + \beta_6 I_{ijt} + \alpha_i + \varepsilon_{it} \quad (4)$$

4.4 Region-specific analyses

The last two analyses that will be done are region specific. The dataset will be split up into two: one dataset which only includes data from the Randstad, and one dataset which only includes data from the non-Randstad area. The COROP division on which the dataset will be split up into Randstad and non-Randstad can be found in Table B1. By analysing the differences between the two datasets, it will be analysed whether a difference in relationship of green space with labour productivity is observed between the two regions. Equation 1 and 2 will be run separately for the two new datasets created.

4.5 Multicollinearity checks

To test the reliability of the results, it is important to test whether variables do not have a strong linear relationship, multicollinearity (Alin, 2010). If the tests prove that there is multicollinearity among two or more variables, the results may not be fully reliable and will result in a loss of precision of estimates (Alin, 2010). As variables are interpreted *ceteris paribus*, i.e. the other variables are held constant, the interpretation could in fact be an impossible outcome with multicollinearity, as the other variable e.g. X has a strong linear

relationship with Y and thus will also change (Alin, 2010). A correlation matrix will be used to assess correlation between variables. A correlation of -1 implies that the variables are perfectly negatively associated, +1 implies that the variables are perfectly positively associated. For this study, a correlation of ± 0.7 will be used as threshold for high correlation, as this value is commonly used (Dormann et al., 2012). Additionally, the Variance Inflation Factor (VIF) tests the portion of variance which is due to multicollinearity; if the VIF is high, it could be an indication of multicollinearity (Alin, 2010). Equation 5 shows the calculation of VIF.

$$\text{VIF} = \frac{1}{1 - R^2} \quad (5)$$

A VIF of one indicates no correlation and a VIF above ten suggests a high level of multicollinearity (Shrestha, 2020). Per model, the VIFs will be calculated based on Pooled OLS regression results, as the VIF test cannot be done after a Fixed Effects model regression.

5. Results

In this section the results of all models will be discussed. The multicollinearity checks show high correlation between labour volume and investments in fixed assets for all models, also the Variance Inflation Factors of these two variables are >10 for all models except the manufacturing industry analysis. Labour volume also is highly correlated (>0.7) with recreational surface in all models, except the business services industry analysis. Lastly, in the Randstad region model, the Variance Inflation Factor of recreational surface is also high (>10), implying a high level of multicollinearity. An extensive overview of the multicollinearity checks can be found in Appendix D.

For all models, a Fixed Effects model is used to estimate the relationship between green space and labour productivity, as the Hausman test was significant for all models, implying that the time-invariant characteristics matter. Therefore, a Fixed Effects model will be less biased than a Random Effects model, as a Fixed Effects model eliminates unit heterogeneity.

5.1 General analysis

First, the general analysis will be discussed, the output can be found in Table 7.

Table 7 Results of the general analysis with ln(labour productivity) as the dependent variable

Variables	(1) Model 1	(2) Model 2	(3) Model 3
Ln(green surface)	-0.8476*** (0.2144)	-0.1575 (0.1365)	
Ln(agricultural surface)			-0.1881* (0.0972)
Ln(recreational surface)			0.9958*** (0.1028)
Ln(woodland and nature surface)			-0.0036 (0.0718)
Labour volume		-0.0002 (0.0009)	-0.0013* (0.0008)
Investments in fixed assets		0.0000*** (0.0000)	0.0001*** (0.0000)
Education high		3.2534*** (0.1811)	2.1310*** (0.1912)
Constant	20.3856*** (2.3382)	11.7969*** (1.4856)	5.0543*** (1.2370)
Observations	360	280	280
R-squared	0.0467	0.7499	0.8235
Number of COROPs	40	40	40

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

Model 1 shows the model without any control variables. A 1% increase in green surface is associated with a decrease in labour productivity of 0.85%, *ceteris paribus*. This association is statistically significant at a 1% significance level. However, as no control variables are included in this model, it is expected to be biased. Therefore, control variables are added to Model 2. A 1% increase in green surface is associated with a decrease in labour productivity of 0.16%, *ceteris paribus*. This association is statistically not significant at a 10% significance level, implying that there is no statistically significant association between green surface and labour productivity. Regarding the control variables, a one-unit increase in labour volume correlates with a decrease of labour productivity of 0.02%, calculated with the following formula: $100 * (\exp(-0.0002) - 1)$, *ceteris paribus*. This association is statistically not significant at a 10% significance level. A one-unit increase in investments of fixed assets is associated with a 0.00% increase in labour productivity, *ceteris paribus*. This association is statistically significant at a 1% significance level. A 0.01-unit (one percentage points) increase in the percentage of people in the working population with high obtained education correlates with an increase in labour productivity of 3.31%, *ceteris paribus*. This association is statistically significant at a 1% significance level. The associations of the control variables are as expected, except for the variable labour volume, but this insignificance could be due to the high level of multicollinearity with investments in fixed assets.

As can be seen in Model 3, there is a different relationship with labour productivity observed for each of the three types of soil included in green space. A 1% increase in agricultural surface correlates with a decrease in labour productivity of 0.18%, *ceteris paribus*. This association is statistically significant at a 10% significance level. A 1% increase in recreational surface is associated with an increase in labour productivity of 1.00%, *ceteris paribus*. This association is statistically significant at a 1% significance level. Lastly, a 1% increase in woodland and nature surface correlates with a decrease in labour productivity of 0.00%, *ceteris paribus*. This association is statistically not significant at a 10% significance level, and thus statistically seen there is no significant relationship between woodland and nature surface and labour productivity. The other (control) variables remain significant and keep the same sign, except for the control variable labour volume. A one-unit increase in labour volume correlates with a 0.13% decrease in labour productivity, *ceteris paribus*. This association is statistically significant at a 1% significance level.

5.2 Industry-specific analyses

Next to the general model, three different industries are analysed as well – the business services, personal services, and manufacturing industry – to see whether the relationship between green space and labour productivity differs across industries.

5.2.1 Business services industry analysis

Table 8 shows the results of the business services industry analysis. Again, three different models will be analysed: the first one without control variables, the second one with all controls, and the last model with the three types of green surface.

Table 8 Results of the business services industry analysis with the natural logarithm of labour productivity of the business services industry as the dependent variable

Variables	(1) Model 1	(2) Model 2	(3) Model 3
Ln(green surface)	-0.5458*** (0.1929)	-0.0523 (0.1501)	
Ln(agricultural surface)			-0.1587 (0.1126)
Ln(recreational surface)			0.8673*** (0.1189)
Ln(woodland and nature surface)			0.1889** (0.0837)
Labour volume business services		-0.0007 (0.0013)	-0.0019 (0.0012)
Investments in fixed assets business services		0.0001*** (0.0000)	0.0001*** (0.0000)
Education high		3.2173*** (0.1905)	2.1787*** (0.2169)
Constant	17.3112*** (2.1036)	10.9509*** (1.6476)	4.1600*** (1.4345)
Observations	360	280	280
R-squared	0.0245	0.6607	0.7371
Number of COROPs	40	40	40

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

Model 1 shows that a 1% increase in green surface correlates with a 0.55% decrease in labour productivity of business services, ceteris paribus. This association is statistically significant at a 1% significance level. Again, this model is likely to be biased as there is not controlled for certain variables which are expected to impact labour productivity as well. Therefore, in Model 2 multiple control variables are introduced.

According to the results of Model 2, a 1% increase in green surface correlates with a decrease in labour productivity of the business services industry of 0.05%, *ceteris paribus*. This association is statistically not significant at a 10% significance level, implying that statistically seen there is no association between the two variables. All control variables are statistically significant at a 1% significance level and positive, which is according to expectations from the literature, except the variable labour volume of business services. A one-unit increase in labour volume of business services is associated with a decrease in labour productivity of this industry of 0.07%, calculated with the formula: $100 * (exp(-0.0007) - 1)$, *ceteris paribus*, and is statistically not significant at a 10% significance level. A one-unit increase in investments in fixed assets in this industry is associated with an increase in this industry's labour productivity of 0.01%, *ceteris paribus*, and is statistically significant at a 1% significance level. A 0.01-unit (one percentage points) increase in the percentage of people in the working population with high obtained education correlates with an increase in labour productivity of this industry of 3.27%, *ceteris paribus*. This association is statistically significant at a 1% significance level.

In Model 3, green space is divided into the three different types of soil it consists of. A 1% increase in agricultural surface is associated with a decrease in the business services' labour productivity of 0.16%, *ceteris paribus*, and is statistically not significant at a 10% significance level. A 1% increase in recreational surface is associated with an increase in this industry's labour productivity of 0.87%, *ceteris paribus*. This association is statistically significant at a 1% significance level. A 1% increase in woodland and nature surface correlates with a 0.19% increase in the labour productivity of business services, *ceteris paribus*, and this association is statistically significant at a 5% significance level. The sign and significance level of the control variables stay the same as in Model 2.

5.2.2 Personal services industry analysis

Table 9 shows the output of the regressions of the personal services industry. Again, the three same models are used for these analyses: one without controls, one with controls, and one with the three different types of green space.

Table 9 Results of the personal services industry analysis with the natural logarithm of labour productivity of the personal services industry as the dependent variable

Variables	(1) Model 1	(2) Model 2	(3) Model 3
Ln(green surface)	-0.9119*** (0.2250)	-0.3095** (0.1292)	
Ln(agricultural surface)			-0.3427*** (0.0909)
Ln(recreational surface)			1.0070*** (0.1010)
Ln(woodland and nature surface)			0.0586 (0.0676)
Labour volume personal services		0.0056*** (0.0015)	0.0011 (0.0013)
Investments in fixed assets personal services		-0.0000 (0.0000)	0.0001* (0.0000)
Education high		2.8186*** (0.1925)	1.8652*** (0.1856)
Constant	20.8693*** (2.4530)	12.9550*** (1.4148)	5.6837*** (1.1714)
Observations	360	280	280
R-squared	0.0490	0.7617	0.8369
Number of COROPs	40	40	40

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

In the personal services industry, a 1% increase in green surface is associated with a 0.91% decrease in labour productivity of this industry, *ceteris paribus*. This association is statistically significant at a 1% significance level. When adding control variables in Model 2, this association becomes smaller but remains negative and statistically significant, now at a 5% significance level. A 1% increase in green surface correlates with a decrease in personal services' labour productivity of 0.31%, *ceteris paribus*, and is statistically significant at a 5% significance level. A one-unit increase in labour volume of the personal services industry is associated with a 0.56% increase in that industry's labour productivity, *ceteris paribus*, and is statistically significant at a 1% significance level. Regarding the investments in fixed assets of this industry, a one-unit increase in that is association with a decrease of 0.00% in this industry's labour productivity, *ceteris paribus*. This association is statistically not significant at a 10% significance level. A 0.01-unit (one percentage points) increase in the percentage of people in the working population with high obtained education correlates with an increase in this industry's labour productivity of 2.86%, *ceteris paribus*. This association is statistically significant at a 1% significance level.

When replacing green surface with the three different types of soil of which it consists of, all types of green except woodland and nature surface are statistically significant. A 1% increase in agricultural surface relates to a decrease of 0.34% in labour productivity of personal services, *ceteris paribus*, and is statistically significant at a 1% significance level. A 1% increase in recreational surface relates to an increase in the personal services' labour productivity of 1.01%, *ceteris paribus*, and this association is statistically significant at a 1% significance level. A 1% increase in woodland and nature surface is associated with an increase of 0.06% in this industry's labour productivity, *ceteris paribus*. This association is statistically not significant at a 10% significance level. There is a change observed in the associations between the control variables and the dependent variable. A one-unit increase in labour volume of personal services is associated with an increase in this industry's labour productivity of 0.11%, *ceteris paribus*, but this association is statistically not significant at a 10% significance level. The coefficient of the variable concerning investments in fixed assets for this industry is now statistically significant at a 10% significance level. A one-unit increase in the personal services industry' investments in fixed assets relates to an increase in its industry's labour productivity of 0.01%, *ceteris paribus*. A 0.01-unit (one percentage points) increase in the percentage of employed people with high obtained education is associated with an increase in the personal services' labour productivity of 1.88%, *ceteris paribus*, and this association is statistically significant at a 1% significance level.

5.2.3 Manufacturing industry analysis

The last industry-specific analysis to be discussed is the manufacturing industry. Table 10 shows the output results of the manufacturing industry specific analysis, including the same three models as in the other industry analyses.

Table 10 Results of the manufacturing industry analysis with the natural logarithm of labour productivity of the manufacturing industry as the dependent variable

Variables	(1) Model 1	(2) Model 2	(3) Model 3
Ln(green surface)	-0.8936*** (0.2584)	-0.0805 (0.1897)	
Ln(agricultural surface)			-0.0992 (0.1513)
Ln(recreational surface)			0.9446*** (0.1597)
Ln(woodland and nature surface)			-0.1343 (0.1118)
Labour volume manufacturing		-0.0164*** (0.0046)	-0.0122*** (0.0044)
Investments in fixed assets manufacturing		0.0004*** (0.0001)	0.0003*** (0.0001)
Education high		2.9155*** (0.2629)	1.9167*** (0.3004)
Constant	21.0763*** (2.8178)	11.5484*** (2.0735)	5.9698*** (1.9135)
Observations	360	280	280
R-squared	0.0361	0.6065	0.6576
Number of COROPs	40	40	40

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

In Model 1, only the main dependent and independent variable of interest are included. A 1% increase in green surface correlates with a decrease of 0.89% in labour productivity of the manufacturing industry, *ceteris paribus*, and this association is statistically significant at a 1% significance level. In Model 2, multiple control variables are added. A 1% increase in green surface is associated with a decrease in the industry's labour productivity of 0.08%, *ceteris paribus*. This association is statistically not significant at a 10% significance level. A one-unit increase in labour volume of the manufacturing industry relates to a decrease in the industry's labour productivity of 1.63%, *ceteris paribus*, and this association is statistically significant at a 1% significance level. A one-unit increase in investment in fixed assets of this industry is associated with an increase in labour productivity of manufacturing of 0.04%, *ceteris paribus*, and this association is statistically significant at a 1% significance level. A 0.01-unit increase (increase of one percentage point) in percentage of people of the working population with high obtained education is related with an increase in this industry's labour productivity of 2.96%, *ceteris paribus*. This association is statistically significant at a 1% significance level.

Model 3 shows the associations of the different types of green with labour productivity. Firstly, a 1% increase in agricultural surface relates with a decrease in labour productivity of

the manufacturing industry of 0.10%, *ceteris paribus*. This association is statistically not significant at a 10% significance level, and thus this association is not statistically significantly different from zero. A 1% increase in recreational surface is associated with an increase of 0.94% in labour productivity of this industry, *ceteris paribus*. This association is statistically significant at a 1% significance level. A 1% increase in woodland and nature surface correlates with a decrease in this industry's labour productivity of 0.13%, *ceteris paribus*. This association is statistically not significant at a 10% significance level, thus not statistically significantly different from zero. The control variables keep the same significance levels and sign as in Model 2, only the size of the association decreases in Model 3.

5.2.4 Comparison of the results of the different industry-specific analyses

In this section, the second and third model of the industry-specific – business services, personal services, and manufacturing – models will be compared, to see whether a difference in the relationship of (types of) green space with labour productivity of the different industries is observed.

Table 11 Comparison of the second models of the industry-specific analyses, with the natural logarithm of labour productivity of the industry mentioned in the model names

Variables	(1) Business services	(2) Personal services	(3) Manufacturing
Ln(green surface)	-0.0523 (0.1501)	-0.3095** (0.1292)	-0.0805 (0.1897)
Labour volume of industry	-0.0007 (0.0013)	0.0056*** (0.0015)	-0.0164*** (0.0046)
Investments in fixed assets of industry	0.0001*** (0.0000)	-0.0000 (0.0000)	0.0004*** (0.0001)
Education high	3.2173*** (0.1905)	2.8186*** (0.1925)	2.9155*** (0.2629)
Constant	10.9509*** (1.6476)	12.9550*** (1.4148)	11.5484*** (2.0735)
Observations	280	280	280
R-squared	0.6607	0.7617	0.6065
Number of COROPs	40	40	40

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

As can be seen in Table 11, only for the personal services industry a statistically significant association between green surface and labour productivity is observed. As mentioned earlier, a 1% increase in green surface correlates with a decrease in labour

productivity of the personal services industry of 0.31%, *ceteris paribus*, and this association is statistically significant at a 5% significance level. The associations of green space with labour productivity of the other industries are statistically not significant. These results are not in line with the hypothesis that green surface has a positive association with labour productivity and are not in line with the hypothesis that a larger positive association is expected in the more labour-intensive industries, such as personal services. The largest negative and significant association between green space and labour productivity is found in the personal services industry.

Table 12 Comparison of Model 3 of the three different industries analysed, with the dependent variable being labour productivity of the industry named in the model

Variables	(1) Business services	(2) Personal services	(3) Manufacturing
Ln(agricultural surface)	-0.1587 (0.1126)	-0.3427*** (0.0909)	-0.0992 (0.1513)
Ln(recreational surface)	0.8673*** (0.1189)	1.0070*** (0.1010)	0.9446*** (0.1597)
Ln(woodland and nature surface)	0.1889** (0.0837)	0.0586 (0.0676)	-0.1343 (0.1118)
Labour volume industry	-0.0019 (0.0012)	0.0011 (0.0013)	-0.0122*** (0.0044)
Investments in fixed assets industry	0.0001*** (0.0000)	0.0001* (0.0000)	0.0003*** (0.0001)
Education high	2.1787*** (0.2169)	1.8652*** (0.1856)	1.9167*** (0.3004)
Constant	4.1600*** (1.4345)	5.6837*** (1.1714)	5.9698*** (1.9135)
Observations	280	280	280
R-squared	0.7371	0.8369	0.6576
Number of COROPs	40	40	40

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

Table 12 shows the third model of all different industry analyses. First, when looking at agricultural green, only for the personal services industry a statistically significant association with labour productivity is found, and this association is negative. The association with labour productivity of the other industries is statistically not significant. Regarding recreational surface, for all three industries a statistically significant association with labour productivity is found at a 1% significance level. The largest positive association is observed in the personal services industry, followed by the manufacturing industry, and the smallest positive significant association is observed in the business services industry. The association

of woodland and nature green with labour productivity is only statistically significant within the business services industry; for the industries the association is statistically significantly not different from zero. This is partially in line with the hypothesis that agricultural land has a negative relationship with labour productivity, as the association is only statistically significant in one of the three industry models. Furthermore, the results of the association of recreational surface and labour productivity are in line with the hypothesis that recreational land has a positive relationship with labour productivity. The largest positive association is observed in the personal services industry, the most labour-intensive industry, which is in line with the hypothesis that (types) of green space has the largest positive association in the labour-intensive industries. However, the association of green space in the most capital-intensive industry, manufacturing, is not the smallest of the three industries, this is not in line with the hypothesis. Lastly, the results of woodland and nature surface are partially in line with the hypothesis that this type of green soil is positively associated with labour productivity, as the association with labour productivity is only statistically significant for the business services industry. However, the largest positive significant association is found the industry that is the second most labour-intensive of the three, which is not in line with the hypothesis that the largest positive association is expected in the most labour-intensive industry. In the personal services industry, no statistically significant association is found on woodland and nature surface with labour productivity.

5.3 Region-specific analyses

Lastly, region-specific analyses are performed to see whether the relationship between (types) of green space and labour productivity differs across regions. The dataset was split up into two subsamples: one subsample only containing COROP areas in the Randstad, and one subsample with the COROP areas outside of the Randstad.

5.3.1 Randstad region analysis

Table 13 shows the output of the regression analyses on the COROP areas located in the Randstad.

Table 13 Results of the Randstad region analysis with the natural logarithm of labour productivity as the dependent variable

Variables	(1) Model 1	(2) Model 2	(3) Model 3
Ln(green surface)	-0.8983*** (0.3134)	-0.1533 (0.1600)	
Ln(agricultural surface)			0.0746 (0.1222)
Ln(recreational surface)			0.6082*** (0.1592)
Ln(woodland and nature surface)			-0.3545*** (0.1246)
Labour volume		-0.0004 (0.0010)	-0.0007 (0.0009)
Investments in fixed assets		0.0000* (0.0000)	0.0000** (0.0000)
Education high		3.1390*** (0.3067)	2.5905*** (0.3196)
Constant	20.3065*** (3.1638)	11.6334*** (1.6240)	7.8428*** (1.8633)
Observations	99	77	77
R-squared	0.0863	0.8169	0.8693
Number of COROPs	11	11	11

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

The first models shows that a 1% increase in green surface is associated with a decrease of 0.90% in labour productivity in the Randstad, *ceteris paribus*. This association is statistically significant at a 1% significance level. As this model is likely to be biased as there are no control variables included, in Model 2 and 3 control variables are added. In Model 2, the association between green surface and labour productivity is statistically not significant at a 10%, and thus the association is statistically significantly not different from zero. A 1% increase in green space correlates with a decrease in labour productivity of 0.15% in the Randstad, *ceteris paribus*. Regarding the control variables, a one-unit increase in labour volume in the Randstad is associated with a decrease in this region's labour productivity with 0.04%, *ceteris paribus*. This association is statistically not significant at a 10% significance level. A one-unit increase in investments in fixed assets correlates with an increase in labour productivity of 0.00%, *ceteris paribus*, and is statistically significant at a 10% significance level. A 0.01-unit increase (a one percentage point increase) in percentage of working population with high obtained education is associated with an increase in labour productivity of 3.19% in the Randstad, *ceteris paribus*. This association is statistically significant at a 1% significance level. Model 3 includes the different types of green soils. A 1% increase in agricultural surface is associated with a 0.07%

increase in labour productivity in the Randstad, *ceteris paribus*. This association is statistically not significant at a 10% significance level. A 1% increase in recreational surface correlates with a 0.61% increase in this region's labour productivity, *ceteris paribus*. This association is statistically significant at a 1% significance level. A 1% increase in woodland and nature surface relates to a 0.35% decrease in labour productivity in the Randstad, *ceteris paribus*. This association is statistically significant at a 1% significance level. The control variables keep the same sign and significance level and only the size of the association decreases, except for the variable investments in fixed assets for which the association is statistically significant at a 5% significance level and no difference in association size is observed.

5.3.2 Non-Randstad region analysis

The other region-specific analyses is on the COROP areas which are located outside of the Randstad, the non-Randstad region, and thus the COROP areas left in the dataset if the Randstad area is dropped out. Table 14 shows the output of the non-Randstad region regressions.

Table 14 Results of the non-Randstad region analysis with the natural logarithm of labour productivity as the dependent variable

Variables	(1) Model 1	(2) Model 2	(3) Model 3
Ln(green surface)	-0.8112*** (0.2872)	-0.1709 (0.1972)	
Ln(agricultural surface)			-0.4236*** (0.1374)
Ln(recreational surface)			1.0457*** (0.1368)
Ln(woodland and nature surface)			0.0876 (0.0848)
Labour volume		0.0017 (0.0014)	-0.0006 (0.0012)
Investments in fixed assets		0.0001*** (0.0000)	0.0001*** (0.0000)
Education high		2.9536*** (0.2206)	1.9050*** (0.2267)
Constant	20.2025*** (3.2195)	11.8210*** (2.2030)	6.4162*** (1.6100)
Observations	261	203	203
R-squared	0.0334	0.7597	0.8329
Number of COROPs	29	29	29

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

In Model 1, no control variables are added. A 1% increase in green surface correlates with a 0.81% decrease in labour productivity in the non-Randstad region, *ceteris paribus*. This association is statistically significant at a 1% significance level. When adding control variables in Model 2, the association of green surface becomes statistically insignificant at a 10% significance level. A 1% increase in green surfaces is associated with a 0.17% decrease in this region's labour productivity, *ceteris paribus*. This association is statistically not significant at a 10% significance level. Regarding the control variables, a one-unit increase in labour volume relates to a 0.17% increase in labour productivity in the non-Randstad region, *ceteris paribus*, and is statistically not significant at a 10% significance level. A one-unit increase in investments in fixed assets correlates with a 0.01% increase in labour productivity in the non-Randstad region, *ceteris paribus*. This association is statistically significant at a 1% significance level. A 0.01-unit (a one percentage point) increase in the percentage of people in the working population with high obtained education correlates with a 3.00% increase in this region's labour productivity, *ceteris paribus*. This association is statistically significant at a 1% significance level.

In Model 3, the three different types of green soil are included. A 1% increase in agricultural surface correlates with a 0.42% decrease in labour productivity in the non-Randstad region, *ceteris paribus*, and is statistically significant at a 1% significance level. A 1% increase in recreational surface relates to a 1.05% increase in labour productivity in this region, *ceteris paribus*. This association is statistically significant at a 1% significance level. A 1% increase in woodland and nature surface is associated with a 0.09% increase in labour productivity in the non-Randstad region, *ceteris paribus*. This association is statistically not significant at a 10% significance level, and thus the association is statistically significantly not different from zero. Regarding the control variables, a one-unit increase in labour volume correlates with a 0.06% decrease in this region's labour productivity, *ceteris paribus*, and is statistically not significant at a 10% significance level. A one-unit increase in investments in fixed assets is associated with a 0.01% increase in labour productivity in the non-Randstad area, *ceteris paribus*, and is statistically significant at a 1% significance level. Lastly, a 0.01-unit (one percentage point) increase in percentage of working people with high obtained education is associated with an increase of 1.92% in this region's labour productivity, *ceteris paribus*. This association is statistically significant at a 1% significance level. The coefficients of the control variables are as expected, except for the insignificant variable labour volume,

but this could be due to the earlier mentioned high level of multicollinearity with investments in fixed assets.

5.3.3 Comparison of the results of the different region-specific analyses

In this section, a comparison of the results of the two region-specific analyses will be discussed. Table 15 shows the second model of the two region-specific models.

Table 15 Comparison of the second models of the region-specific analyses, with the natural logarithm of labour productivity from the region as dependent variable

Variables	(1) Randstad	(2) Non-Randstad
Ln(green surface)	-0.1533 (0.1600)	-0.1709 (0.1972)
Labour volume of industry	-0.0004 (0.0010)	0.0017 (0.0014)
Investments in fixed assets	0.0000* (0.0000)	0.0001*** (0.0000)
Education high	3.1390*** (0.3067)	2.9536*** (0.2206)
Constant	11.6334*** (1.6240)	11.8210*** (2.2030)
Observations	77	203
R-squared	0.8169	0.7597
Number of COROPs	11	29

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

As can be seen in Table 15, no difference is observed in the association of green space with labour productivity between the two regions; the association of green surface is statistically not significant at a 10% significance level in both region-specific models, and thus the association in both models is not statistically different from zero. This is not in line with the hypothesis that in the Randstad region a larger positive significant association with labour productivity was expected. Table 16 shows the output of the third models of both regions.

Table 16 Comparison of the third models of the region-specific analyses, with the natural logarithm of labour productivity as dependent variable

Variables	(1) Randstad	(2) Non-Randstad
Ln(agricultural surface)	0.0746 (0.1222)	-0.4236*** (0.1374)
Ln(recreational surface)	0.6082*** (0.1592)	1.0457*** (0.1368)
Ln(woodland and nature surface)	-0.3545*** (0.1246)	0.0876 (0.0848)
Labour volume	-0.0007 (0.0009)	-0.0006 (0.0012)
Investments in fixed assets	0.0000** (0.0000)	0.0001*** (0.0000)
Education high	2.5905*** (0.3196)	1.9050*** (0.2267)
Constant	7.8428*** (1.8633)	6.4162*** (1.6100)
Observations	77	203
R-squared	0.8693	0.8329
Number of COROPs	11	29

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

Lastly, when looking at the different types of green soils, a difference in associations is observed between the two regions. Agricultural surface has a statistically significant negative association with labour productivity in the non-Randstad area, whereas it has no significant association with labour productivity in the Randstad area. Furthermore, recreational surface has a larger significant positive association with labour productivity in the non-Randstad area, which is contrary to the hypothesis that green areas have a larger positive association in the Randstad area. Lastly, woodland and nature surface has a significant negative association with labour productivity in the Randstad, whereas it has no statistically significant association in the non-Randstad area. None of these results are in line with the hypothesis that the association of (types) of green space and labour productivity is larger and positive in the Randstad compared to the non-Randstad area.

6. Discussion

The findings of this research are partially in line with the hypotheses. In this section possible explanations of unexpected findings, research limitations, implications of the results, and recommendations for further research are discussed.

6.1 Hypotheses

The first hypothesis that an increase in green surface is expected to increase the aggregated labour productivity in the Netherlands, was found not to be true in this study: only statistically significant negative or statistically insignificant associations were found for all models. This unexpected outcome might be explained by agricultural surface. Out of the three types of soil included in green space, agricultural land takes the largest part in the Netherlands, according to Table C1. This could be a potential explanation for aggregated green space to have a statistically insignificant or even statistically negative effect; agricultural activities tend to have a relatively lower labour productivity compared to the country average labour productivity according to Dieppe & Matsuoka (2021) and also according to the descriptive statistics in Table C1. If we take green space, including agricultural land, it might implicate that a higher number of green space goes in hand with higher agricultural surface, and thus an average lower labour productivity. Aggregate green space might include three types of soil with different directions of associations, and combining these together could lead to insignificant results. Also to test this hypothesis, it was decided to add a third model to analyse the three different types of green soil of which green space consists of.

The second hypothesis on the effects of the three different types of soil, where a negative relationship is expected between agricultural land and productivity, and a positive relationship is expected between the other two types of green soil – recreational, woodland, and nature surface – and labour productivity is partially found to be true. Agricultural surface is found to have a statistically insignificant association with labour productivity in the business services, manufacturing industry, and Randstad-region models. As the business services and manufacturing industry are relatively less labour intensive, compared to the personal services industry, it could be that therefore labour productivity is less sensitive to changes in types of green, and therefore no significant association was found. The more capital-intensive industries obtain higher labour productivity due to their capital intensity (Guest, 2011). The labour productivity of the more capital-intensive industries could therefore less dependent on changes in human capital. Thus, if we expect green soils to have a relationship with labour productivity

with an intermediate effect such as human capital, it might explain why the business services and manufacturing industry are less sensitive to changes in (types) of green soils. Furthermore, the number of observations in the Randstad model is quite low, and it could be that therefore no significant association was found. The agricultural surface associations were as expected in the other models. The relationship between recreational land and labour productivity was significantly positive in all models, which is in line with the second hypothesis. Regarding woodland and nature surface, only for the business services industry a significantly positive association was found. For the Randstad analysis a negative association was found, and for the other models no significant association was found. The insignificant relationship of woodland and nature surface with labour productivity might be explained by the characteristics of the soil type. For example, Zijlema et al. (2017) found that cognitive abilities depend on the nearness of natural areas. Dunes for example, which are part of this type of soil, are often located a bit further away and could be less accessible, implying that the nearness of such soils is already lower than other types of green soils, and therefore limiting potential exposure. Furthermore, from the correlation matrices in Appendix D it can be observed that labour volume often is least correlated with woodland and nature, compared to the other types of green soils, which might indicate that the areas with high labour volume are not closely located to the areas with large amounts of woodland and nature surface.

The third hypothesis that a larger positive significant effect is expected in labour-intensive industries than capital-intensive industries is also partially found to be true. The personal services industry is defined as the most labour-intensive industry, and the manufacturing industry is defined as the most capital-intensive industry. In the results, it is found that the change in the personal services industry' labour productivity is the largest of all industries when a change green space is observed. Aggregated green surface is found to have a statistically insignificant relationship with labour productivity, except for personal services industry: that relationship is statistically significantly negative. This is contrary to the expected effect but could be due to the earlier mentioned explanation that a large part of green surface is agricultural surface, which is expected to have a negative effect on labour productivity. When looking at the models with the three different types of soils, it is found that for all types of soils, except woodland and nature surface, the association with labour productivity is the largest for the personal services industry and the smallest for the manufacturing industry. This is in line with the hypothesis. Guest (2011) also mentions that in the capital-intensive industries labour productivity is largely determined by its capital intensity. This could explain why the

relationship of (types) of green spaces with labour productivity of the more capital-intensive industries changes less or not compared to labour-intensive industries when a change in green soils is observed. A possible explanation for woodland and nature surface being insignificant for personal services industry might be the type of greenery, as explained before.

Lastly, the fourth hypothesis that (types of) green spaces are expected to have a larger positive association with labour productivity in the Randstad compared to the non-Randstad is found not to be true in all models. For aggregate green surface no difference in association is found as that variable is statistically insignificant for both regions. The insignificance could be due to the contrary differences in associations between the three types of green soil which might make the association of aggregate green statistically insignificant, as discussed earlier. Agricultural surface is only negatively associated with labour productivity in the non-Randstad area, and insignificant for the Randstad. An explanation of the insignificant association could be the relatively low surface and low within variance of agricultural land in the Randstad compared to the non-Randstad area. The association of recreational surface with labour productivity is found to be statistically significantly positive in both regions, and larger in the non-Randstad region. As recreational surface is found to be scarcer in the non-Randstad, see Table C1, it could be that with an increase in recreational surface the marginal benefits are higher for the non-Randstad area. Lastly, woodland and nature surface is found to have a significantly negative association with labour productivity in the Randstad, compared to no significant effect in the non-Randstad area, which is also not in line with the hypothesis. In the data it was found that over the years woodland and nature surface in the Randstad did not increase as much compared to the non-Randstad, and in some years, the surface even decreased compared to the year before, while labour productivity was increasing. According to CBS (2024), woodland and nature surface has been decreasing between 2013 and 2022 to make space for buildings, including offices. The decrease of woodland and nature surface was the highest in the province South-Holland, part of the Randstad, while the non-Randstad area observed increases in woodland and nature (CBS, 2024). This could explain the negative association between woodland and nature surface with labour productivity in the Randstad.

6.2 Research limitations

While this research provides insights into the relationship between green surface and labour productivity, it is important to acknowledge certain limitations that may impact the

reliability of the findings. During this study, multiple assumptions and decisions are made which perhaps impact this research' validity.

First of all, the use of COROP areas will be discussed. There are two main reasons why COROP areas were used as location indicator: no differences between work- and living locations and the lack of data as the smallest area on which all datasets are available is on COROP level. By using COROP areas in this study, there is accounted for differences in work- and living areas and thus assumed that people live and work in the same COROP area, as Rijksinstituut voor Volksgezondheid en Milieu (n.d.) made these areas based on commuter flows. However, this assumption seems to be unrealistic. Inhabitants of regions seem to commute between COROP regions, according to the dataset on jobs of employees in which the number of employees is above zero between different work- and living COROP area combinations (CBS, 2016). This could imply that people are exposed to greenery of multiple COROPs, for example when the home location (also the location from which people work if they decide to work from home) is in COROP 01, and the work location is in COROP 02, making the results less reliable. What if people from a certain COROP area on average travel more often to another COROP for work compared to another COROP, this could make the results ambiguous, because by the change of which COROP greenery is the increase in labour productivity caused? To account for this unknown effect, it might be interesting to analyse only data from the people that work from home, to make sure the assumption of people working and living in the same COROP is validated, and to isolate the effect of greenery on labour productivity purely within one COROP. Furthermore, the smallest scale on which all datasets used had information available was on COROP level. A smaller scale might be more desirable, as most literature on the relationship between green and health was on neighbourhood or city-level, however this smaller scale data was not available. This difference in scale might imply that the results should be interpreted rather as capturing the relationship as having a higher chance of being exposed to greenery, instead of physically seeing greenery, as an increase in greenery on such large scale does not have to imply that all people see this green surface. However, this research therefore might give insights to a whole new level of the effect of green: a larger scale effect which has not been researched till now.

Another limitation is that there might be a sorting effect: people who appreciate green more live and work in the greener areas, or another potential sorting effect is that it might be that the more productive people work in greener areas or the other way around. For example, Braçe et al. (2020) found that the positive effect of greenery on (mental) health is only

statistically significant for people who appreciate green views. To mitigate this potential limitation, a Fixed Effects model is used. With the use of this model there is partially taken care of the sorting effect: the time-invariant characteristics are eliminated. However, the characteristics that do vary over time that could introduce a sorting effect are not accounted for. During the process no strong and valid instrumental variable was found, as the potential instruments did not only have an effect on green space, but labour productivity as well or they were not available for all years and all COROP areas analysed. With the use of an instrumental variable the results will be less biased.

Lastly, presence of multicollinearity is high likely. Regarding the variables labour volume and investments in fixed, the level of multicollinearity is high as the Variance Inflation Factor is above the threshold (>10), indicating that the variance of the coefficients has increased due to multicollinearity. The correlation matrices also show high levels correlation (>0.7) for these two variables. As there are only three control variables, no other potential control variables were available due to lack of data, and according to the literature these variables could explain a large part of labour productivity, it was chosen to leave these variables in the regression models, however, as there is multicollinearity, this brings some limitations. Consequences of leaving in multicollinear variables are that the standard errors increase, potentially resulting in insignificant coefficients, while these variables are still important. Furthermore, it is difficult to determine what the individual association is of one variable, as they are multicollinear; it is unrealistic that one of these variables will stay constant while the other one changes in value. Therefore, the results of the regressions are less reliable due to this multicollinearity.

6.3 Implications and recommendations

Due to the potential limitations from the previous section, it is recommended to interpret the results on the relationship between green space and labour productivity more as an association rather than a true causal effect. The following suggestions for further research could be implemented to account for these limitations.

As discussed earlier, there could be a potential sorting effect: people who appreciate green more, live and work more in greener areas. If this is the case, the results are less reliable, as the relationship with green spaces might not be applicable to all people. For further research it is therefore recommended to make use of an instrumental variable, if the data is available to do so. During this research, there was no suitable instrumental variable found, but if there is a

strong and valid instrumental variable available which only has an effect on the main independent variable green space, it would be suggested to implement this in further research.

Furthermore, as the data on soil was only available till 2017, the effect of COVID-19 could not be analysed. For further research, it could be interesting to see how the effect of green space changed after COVID-19 and to analyse more recent data. During the COVID-19 period, people's mental health has become worse (Rijksinstituut voor Volksgezondheid en Milieu, 2023). During and after COVID-19, there has been a trend in walking more often and an increase in working from home (Ministerie van Infrastructuur en Waterstaat, 2021; RTV Utrecht, 2022). Due to this, the relationship between green space and labour productivity might have changed compared to the time before COVID-19. A Difference-in-Difference model could be used to do this research. Furthermore, it might be interesting to look at the difference in effect between different types of working people, such as employees and self-employed people. A difference between these two groups might be expected as self-employed people could have more variety in working locations and having less strict hours.

Lastly, as it is recommended to interpret this more as an association rather than a true causal effect and because there is multicollinearity in the data, it is recommended to do further research in this topic to truly estimate the causal effect before implementing the findings of this study into policy making. If data becomes available on other control variables, which are not used in this research, multicollinearity might be prevented, as more data is available, making the results more reliable. However, this research could be used as base research, as a starting point for new studies, as it shows the gaps in existing literature and what the difficulties are in identifying a true causal effect between green spaces and labour productivity.

7. Conclusion

In this study, the relationship between (types of) green space and labour productivity in the Netherlands has been analysed. Based on the results, it can be concluded that no significant relationship has been found between aggregate green space and labour productivity. However, when looking into different types of green spaces, statistically significant results have been found. An increase in agricultural surface significantly correlates with a decrease in labour productivity in the Netherlands. Contrary to this, an increase in recreational surface is significantly associated with an increase in labour productivity in the Netherlands. No significant association has been found for woodland and nature surface on labour productivity.

Next to totals of labour productivity in the Netherlands, three different industries have been analysed as well. Only for the personal services industry a significant negative association between green surface and labour productivity has been found, no significant associations were found for the other two industries, business services and manufacturing. After looking into the different types of green soil, agricultural surface significantly correlates with a decrease in labour productivity of the personal services industry, and no significant relationships have been found for the other two industries. The relationship between recreational surface and labour productivity was found to be significantly positive for all industries, with the largest association in the most labour-intensive industry: personal services, followed by the manufacturing industry. The smallest positive association was found for the business services industry. For woodland and nature surface, only a significant positive association was found with labour productivity for the business services industry. No significant association between woodland and nature surface and labour productivity was found for the personal services and manufacturing industry.

Lastly, regional analyses have been done as well: a comparison of the associations found between the Randstad and non-Randstad area. No difference in relationship between green space and labour productivity has been found; aggregate green space was not significant for both regions. The association of agricultural surface and labour productivity was found to be significantly negative for the non-Randstad area, and no significant association was found for the Randstad. Recreational surface was significantly positive for both regions, where a larger positive association with labour productivity was found for the non-Randstad region. The association between woodland and nature surface and labour productivity was found to be statistically insignificant for the non-Randstad region, and significantly negative for the Randstad.

8. Reference List

- Aan de Brugh, M. (2024, May 15). Hoe groen moet de stad zijn? Zie maar eens te bewijzen dat een groene omgeving goed is voor mensen. *NRC*. <https://www.nrc.nl/nieuws/2024/05/15/hoegroenmoetdestadzijnziemaar eens te bewijzen dat een groene omgeving goed is voor mensen-a4198900#:~:text=Mensen%20die%20in%20een%20groenere,hebben%20gemiddeld%20een%20hoger%20geboortegewicht.>
- Adedeji, O., & Campbell, O. (2013). The Role of Higher Education in Human Capital Development. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.2380878>
- Akpınar, A., Barbosa-Leiker, C., & Brooks, K. R. (2016). Does green space matter? Exploring relationships between green space type and health indicators. *Urban Forestry & Urban Greening*, 20, 407–418. <https://doi.org/10.1016/j.ufug.2016.10.013>
- Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews. Computational Statistics*, 2(3), 370–374. <https://doi.org/10.1002/wics.84>
- Amini, S., Delgado, M. S., Henderson, D. J., & Parmeter, C. F. (2012). Fixed vs Random: The Hausman Test Four Decades Later. In *Advances in econometrics* (pp. 479–513). [https://doi.org/10.1108/s0731-9053\(2012\)0000029021](https://doi.org/10.1108/s0731-9053(2012)0000029021)
- Anos-Casero, P., & Udomsaph, C. (2009). What drives firm productivity growth? In *World Bank policy research working paper*. <https://doi.org/10.1596/1813-9450-4841>
- Berman, M. G., Jonides, J., & Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychological Science*, 19(12), 1207–1212. <https://doi.org/10.1111/j.1467-9280.2008.02225.x>
- Berto, R. (2005). Exposure to restorative environments helps restore attentional capacity. *Journal of Environmental Psychology*, 25(3), 249–259. <https://doi.org/10.1016/j.jenvp.2005.07.001>
- Bloom, D. E., & Canning, D. (2003). Health as Human Capital and its Impact on Economic Performance. *The Geneva Papers on Risk and Insurance*, 28(2), 304–315. <https://doi.org/10.1111/1468-0440.00225>

- Braçe, O., Garrido-Cumbrera, M., Foley, R., Correa-Fernández, J., Suárez-Cáceres, G. P., & Laforteza, R. (2020). Is a View of Green Spaces from Home Associated with a Lower Risk of Anxiety and Depression? *International Journal of Environmental Research and Public Health*, *17*(19), 7014. <https://doi.org/10.3390/ijerph17197014>
- Bratman, G. N., Daily, G. C., Levy, B. J., & Gross, J. J. (2015). The benefits of nature experience: Improved affect and cognition. *Landscape and Urban Planning*, *138*, 41–50. <https://doi.org/10.1016/j.landurbplan.2015.02.005>
- Bringslimark, T., Hartig, T., & Patil, G. G. (2007). Psychological Benefits of Indoor Plants in Workplaces: Putting Experimental Results into Context. *HortScience*, *42*(3), 581–587. <https://doi.org/10.21273/hortsci.42.3.581>
- Buckley, R. C., & Chauvenet, A. L. (2022). Economic value of nature via healthcare savings and productivity increases. *Biological Conservation*, *272*, 109665. <https://doi.org/10.1016/j.biocon.2022.109665>
- Cardinali, M., Beenackers, M. A., Fleury-Bahi, G., Bodénan, P., Petrova, M., Van Timmeren, A., & Pottgiesser, U. (2024). Examining Green Space Characteristics for Social Cohesion and Mental Health Outcomes: A Sensitivity Analysis in four European Cities. *Urban Forestry & Urban Greening*, 128230. <https://doi.org/10.1016/j.ufug.2024.128230>
- CBS. (n.d.). *COROP-gebied*. <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/corop-gebied#:~:text=Een%20cluster%20van%20%C3%A9%20n%20of,door%20de%20CO%20B6rdinatiecommissie%20Regionaal%20OnderzoeksProgramma>.
- CBS. (n.d.-b). *Waar wonen de meeste hoogopgeleiden?* <https://www.cbs.nl/nl-nl/dossier/dossier-verstedelijking/waar-wonen-de-meeste-hoogopgeleiden>
- CBS. (2014). *Beroepsbevolking; regio's 1996-2013* [Dataset]. <https://opendata.cbs.nl/statline/#!/CBS/nl/dataset/71887ned/table?ts=1716819024320>

- CBS. (2016). *Banen van werknemers; geslacht, leeftijd, woon- en werkregio's (2006-2014)* [Dataset].
<https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81252NED/table?ts=1708961673271>
- CBS. (2022, August 12). *De Randstad*. <https://www.cbs.nl/nl-nl/nieuws/2022/32/trek-uit-de-randstad-blijft-toenemen/de-randstad>
- CBS. (2023a). *Bodemgebruik; uitgebreide gebruiksvorm, per gemeente* [Dataset].
<https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70262ned/table?fromstatweb>
- CBS. (2023b, September 28). Arbeidsproductiviteit verder toegenomen in 2022. *CBS*.
<https://www.cbs.nl/nl-nl/nieuws/2023/39/arbeidsproductiviteit-verder-toegenomen-in-2022>
- CBS. (2023c). *Investerings in vaste activa; bedrijfstak en regio, nationale rekeningen* [Dataset].
<https://opendata.cbs.nl/StatLine/#/CBS/nl/dataset/84440NED/table?ts=1714924718636>
- CBS. (2023d). *Productieproces; bedrijfstak en regio, nationale rekeningen* [Dataset].
<https://opendata.cbs.nl/#/CBS/nl/dataset/84419NED/table>
- CBS. (2024a, May 2). Minder natuur- en bosgebied in Nederland. *Centraal Bureau Voor De Statistiek*. <https://www.cbs.nl/nl-nl/nieuws/2024/18/minder-natuur-en-bosgebied-in-nederland>
- Chang, C., & Chen, P. (2005). Human response to window views and indoor plants in the workplace. *HortScience*, 40(5), 1354–1359. <https://doi.org/10.21273/hortsci.40.5.1354>
- Chansarn, S. (2010). Labor Productivity Growth, Education, Health and Technological Progress: A Cross-Country Analysis. *Economic Analysis & Policy*, 40(2).
- Choi, J., Park, S., Jung, S., Lee, J., Son, K., An, Y., & Lee, S. (2016). Physiological and psychological responses of humans to the index of greenness of an interior space. *Complementary Therapies in Medicine*, 28, 37–43.
<https://doi.org/10.1016/j.ctim.2016.08.002>

- Chung, W. K., Chau, C. K., Masullo, M., & Pascale, A. (2019). Modelling perceived oppressiveness and noise annoyance responses to window views of densely packed residential high-rise environments. *Building and Environment*, *157*, 127–138. <https://doi.org/10.1016/j.buildenv.2019.04.042>
- Compendium voor de Leefomgeving. (2023, June 8). Veranderingen bodemgebruik, 1996 - 2017. <https://www.clo.nl/indicatoren/nl006012-veranderingen-bodemgebruik-1996-2017>
- Dieppe, A., & Matsuoka, H. (2021). Sectoral sources of productivity growth. In *The World Bank eBooks* (pp. 357–389). https://doi.org/10.1596/978-1-4648-1608-6_ch7
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2012). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- European Parliament. (2023, July). *Mental health in the EU*. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751416/EPRS_BRI\(2023\)751416_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751416/EPRS_BRI(2023)751416_EN.pdf)
- Fan, Y., Das, K., & Chen, Q. (2011). Neighborhood green, social support, physical activity, and stress: Assessing the cumulative impact. *Health & Place*, *17*(6), 1202–1211. <https://doi.org/10.1016/j.healthplace.2011.08.008>
- Fedulova, I., Voronkova, O. Y., Zhuravlev, P., Gerasimova, E. A., Glyzina, M. P., & Alekhina, N. A. (2019). Labor productivity and its role in the sustainable development of economy: on the example of a region. *Entrepreneurship and Sustainability Issues*, *7*(2), 1059–1073. [https://doi.org/10.9770/jesi.2019.7.2\(19\)](https://doi.org/10.9770/jesi.2019.7.2(19))
- Fjeld, T., Veiersted, B., Sandvik, L., Riise, G., & Levy, F. (1998). The Effect of Indoor Foliage Plants on Health and Discomfort Symptoms among Office Workers. *Indoor + Built Environment*, *7*(4), 204–209. <https://doi.org/10.1177/1420326x9800700404>

- Fried, H. O., Knox Lovell, C. A., & Schmidt, S. S. (1993). The measurement of productive efficiency and productivity change. *Oxford University Press*.
<https://doi.org/10.1093/oso/9780195072181.001.0001>
- Guest, R. (2011). Population ageing, capital intensity and labour productivity. *Pacific Economic Review*, *16*(3), 371–388. <https://doi.org/10.1111/j.1468-0106.2011.00553.x>
- Hartig, T., Evans, G. W., Jamner, L. D., Davis, D., & Gärling, T. (2003). Tracking restoration in natural and urban field settings. *Journal of Environmental Psychology*, *23*(2), 109–123. [https://doi.org/10.1016/s0272-4944\(02\)00109-3](https://doi.org/10.1016/s0272-4944(02)00109-3)
- Horr, Y. A., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., & Elsarrag, E. (2016). Occupant productivity and office indoor environment quality: A review of the literature. *Building and Environment*, *105*, 369–389.
<https://doi.org/10.1016/j.buildenv.2016.06.001>
- Hsiao, C. (2003). *Analysis of panel data*. <https://doi.org/10.1017/cbo9780511754203>
- Huffman, W. E. (2001). Chapter 7 Human capital: Education and agriculture. In *Handbook of agricultural economics* (pp. 333–381). [https://doi.org/10.1016/s1574-0072\(01\)10010-1](https://doi.org/10.1016/s1574-0072(01)10010-1)
- Ikei, H., Song, C., Igarashi, M., Namekawa, T., & Miyazaki, Y. (2014). Physiological and psychological relaxing effects of visual stimulation with foliage plants in high school students. *Advances in Horticultural Science*, *28*(2), 111–116.
<https://doi.org/10.13128/ahs-22823>
- Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, *15*(3), 169–182. [https://doi.org/10.1016/0272-4944\(95\)90001-2](https://doi.org/10.1016/0272-4944(95)90001-2)
- Korkmaz, S., & Korkmaz, O. (2017). The Relationship between Labor Productivity and Economic Growth in OECD Countries. *International Journal of Economics and Finance*, *9*(5), 71. <https://doi.org/10.5539/ijef.v9n5p71>

- Lohr, V. I., Pearson-Mims, C. H., & Goodwin, G. K. (1996). Interior plants may improve worker productivity and reduce stress in a windowless environment. *Journal of Environmental Horticulture*, *14*(2), 97–100. <https://doi.org/10.24266/0738-2898-14.2.97>
- Lottrup, L., Grahn, P., & Stigsdotter, U. K. (2013). Workplace greenery and perceived level of stress: Benefits of access to a green outdoor environment at the workplace. *Landscape and Urban Planning*, *110*, 5–11. <https://doi.org/10.1016/j.landurbplan.2012.09.002>
- Lovell, C. K. (1993). Production frontiers and productive efficiency. *The measurement of productive efficiency: techniques and applications*, *3*, 67.
- Luo, L., & Jiang, B. (2022). From oppressiveness to stress: A development of Stress Reduction Theory in the context of contemporary high-density city. *Journal of Environmental Psychology*, *84*, 101883. <https://doi.org/10.1016/j.jenvp.2022.101883>
- 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*, *60*(7), 587–592. <https://doi.org/10.1136/jech.2005.043125>
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, *107*(2), 407–437. <https://doi.org/10.2307/2118477>
- Menard, S. W. (2008). *Handbook of Longitudinal Research: Design, Measurement, and Analysis*. <http://ci.nii.ac.jp/ncid/BA84148212>
- Ministerie van Algemene Zaken. (2023, September 20). *Plannen voor Wonen en bouwen*. Prinsjesdag: Miljoenennota En Rijksbegroting | Rijksoverheid.nl. <https://www.rijksoverheid.nl/onderwerpen/prinsjesdag/wonen-en-bouwen>
- Ministerie van Infrastructuur en Waterstaat. (2021, July 19). *Onderzoek wijst uit: thuiswerken is een blijvertje*. Nieuwsbericht | Rijksoverheid.nl. <https://www.rijksoverheid.nl/actueel/nieuws/2021/07/14/onderzoek-wijst-uit-thuiswerken-is-een-blijvertje>

- Mouissie, S., & Kraniotis, L. (2023, November 11). Schreeuwend tekort aan woningen en hoge huizenprijzen: hoe is het zo gekomen? *NOS*. <https://nos.nl/collectie/13960/artikel/2497415-schreeuwend-tekort-aan-woningen-en-hoge-huizenprijzen-hoe-is-het-zo-gekomen>
- Newsham, G. R., Birt, B. J., Arsenault, C., Thompson, A., Veitch, J. A., Mancini, S., Galasiu, A. D., Gover, B. N., Macdonald, I., & Burns, G. J. (2013). Do ‘green’ buildings have better indoor environments? New evidence. *Building Research and Information*, *41*(4), 415–434. <https://doi.org/10.1080/09613218.2013.789951>
- NOS. (2021, March 6). Politieke partijen willen meer huizen bouwen, maar waar? *NOS*. <https://nos.nl/artikel/2371523-politieke-partijen-willen-meer-huizen-bouwen-maar-waar>
- Öhrström, E. (2004). Longitudinal surveys on effects of changes in road traffic noise—annoyance, activity disturbances, and psycho-social well-being. *The Journal of the Acoustical Society of America*, *115*(2), 719–729. <https://doi.org/10.1121/1.1639333>
- Puga, D. (2010). The Magnitude and Causes of Agglomeration Economies. *Journal of Regional Science*, *50*(1), 203–219. <https://doi.org/10.1111/j.1467-9787.2009.00657.x>
- Raad voor de Leefomgeving en Infrastructuur, Raad voor het Openbaar Bestuur, & Raad voor Volksgezondheid & Samenleving. (2023, March). *Adviesrapport Rli, ROB en RVS Elke regio telt! Een nieuwe aanpak van verschillen tussen regio's*. Elke Regio Telt. <https://www.elkeregiotelt.nl/publicaties/rapporten/2023/03/27/adviesrapport-rli-rob-rvs-elke-regio-telt>
- Richardson, E. A., Mitchell, R., Hartig, T., De Vries, S., Astell-Burt, T., & Frumkin, H. (2011). Green cities and health: a question of scale? *Journal of Epidemiology and Community Health*, *66*(2), 160–165. <https://doi.org/10.1136/jech.2011.137240>
- Rijksinstituut voor Volksgezondheid en Milieu. (n.d.). *Regio's en regio-indelingen*. Regiobeeld. <https://www.regiobeeld.nl/regios>
- Rijksinstituut voor Volksgezondheid en Milieu. (2023, April 20). *Invloed van de coronapandemie op de gezondheid en leefstijl van Nederlandse volwassenen*.

<https://www.rivm.nl/publicaties/invloed-corona-epidemie-op-gezondheid-leefstijl-volwassenen#:~:text=Zo%20gaven%20mensen%20aan%20zich,de%204%2C5%20procent>).

RTV Utrecht. (2022, March 27). Wandelen, wandelen en nog eens wandelen in coronatijd: “Met name de ommetjes zijn erg populair geworden.” *RTV Utrecht*. <https://www.rtvutrecht.nl/nieuws/3239320/wandelen-wandelen-en-nog-eens-wandelen-in-coronatijd-met-name-de-ommetjes-zijn-erg-populair-geworden>

Russell, R., Guerry, A. D., Balvanera, P., Gould, R. K., Basurto, X., Chan, K. M. A., Klain, S. C., Levine, J., & Tam, J. (2013). Humans and Nature: How knowing and Experiencing Nature Affect Well-Being. *Annual Review of Environment and Resources*, 38(1), 473–502. <https://doi.org/10.1146/annurev-environ-012312-110838>

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. <https://doi.org/10.12691/ajams-8-2-1>

Taylor, A. F., & Kuo, F. E. (2009). Children with attention deficits concentrate better after walk in the park. *Journal of Attention Disorders*, 12(5), 402–409. <https://doi.org/10.1177/1087054708323000>

Tennessen, C. M., & Cimprich, B. (1995). Views to nature: Effects on attention. *Journal of Environmental Psychology*, 15(1), 77–85. [https://doi.org/10.1016/0272-4944\(95\)90016-0](https://doi.org/10.1016/0272-4944(95)90016-0)

Trpeski, P., Cvetanoska, M., & Kozheski, K. (2019, December 1). Physical capital investments and labour productivity across countries - panel approach. *Repository of UKIM*. <http://hdl.handle.net/20.500.12188/6541>

Ulrich, R. S. (1984). View Through a Window May Influence Recovery from Surgery. *Science*, 224(4647), 420–421. <https://doi.org/10.1126/science.6143402>

Ulrich, R. S. (1991). Effects of interior design on wellness: theory and recent scientific research. *PubMed*, 3, 97–109. <https://pubmed.ncbi.nlm.nih.gov/10123973>

- Weisbrod, B. A. (1962). Education and investment in human capital. *Journal of Political Economy*, 70(5, Part 2), 106–123. <https://doi.org/10.1086/258728>
- Wilson, E. O. (1984). Biophilia. In *Harvard University Press eBooks*. <https://doi.org/10.4159/9780674045231>
- World Health Organization. (2001). *Macroeconomics and Health: Investing in Health for Economic Development*. <https://iris.who.int/bitstream/handle/10665/42435/92415-4550X.pdf?sequence=1>
- Yang, L., Ho, Y. S., Wong, F. K. Y., Chang, K. K. P., Chan, K. C., Wong, M. S., Ho, H. C., Yuen, J. W., Huang, J., & Siu, J. Y. (2020). Neighbourhood green space, perceived stress and sleep quality in an urban population. *Urban Forestry & Urban Greening*, 54, 126763. <https://doi.org/10.1016/j.ufug.2020.126763>
- Yildirim, K., Koyuncu, C., & Koyuncu, J. (2009). Does temperature affect labor productivity: Cross-country evidence. *Applied Econometrics and International Development*, 9(1), 29-39.
- Zijlema, W. L., Triguero-Mas, M., Smith, G., Cirach, M., Martínez, D., Dadvand, P., Gascón, M., Jones, M. V., Gidlow, C., Hurst, G., Masterson, D., Ellis, N., Van Den Berg, M., Maas, J., Van Kamp, I., Van Den Hazel, P., Kruize, H., Nieuwenhuijsen, M., & Júlvez, J. (2017). The relationship between natural outdoor environments and cognitive functioning and its mediators. *Environmental Research*, 155, 268–275. <https://doi.org/10.1016/j.envres.2017.02.017>

Appendix A | Literature on the effect of indoor greenery on (mental) health

Multiple studies have also been conducted on the effects of indoor greenery on health. Newsham et al. (2013) found that green office buildings, including higher window availability, led to better mood and sleep quality. Choi et al. (2016) did experimental research on subjective to indoor greenery and found that already a small portion of indoor foliage can have a calming effect on individuals. Ikei et al. (2014) did research on the heart rate variability by comparing people in a setting with and without indoor greenery. The authors found that plants caused an increase in the autonomic nerve system's stability, implying a relaxation of the nerve system. Bringslimark et al. (2007) found that employees' self-reported sick leave has an association with the level of presence of indoor foliage close to the employees' desks. An analysis was done on cross-sectional survey data of Norwegian employees. According to research of Lohr et al. (1996), individuals' blood pressure is lower in an environment with greenery compared to a room without greenery. The authors conducted this experimental study in a classroom without windows. This research suggests that indoor greenery may lower stress levels (Lohr et al., 1996). Chang and Chen (2005) also did an experimental study and found that anxiety and nervousness levels decreased when people were placed in a room with plants, compared to a room without greenery. Cross-over research of Fjeld et al. (1998) states that company buildings with plants caused a decrease in employee exhaustion complaints. The research included twelve different questions regarding health symptoms distributed among Norwegian employees (Fjeld et al., 1998). The results of the study of Fjeld et al. (1998) include a significant decrease of 23 percent of all symptoms of participants being in an office with greenery compared to an office without. On larger scales than company-level there seems a scarcity in identifying the effect of indoor green. Most research on the effect of indoor green has been done with experimental design or self-reported data, and so far, the quantity of indoor green has not been measured, rather whether there is indoor greenery or not on a small scale.

Appendix B | Data description

Table B1 Overview of codes and corresponding names of COROP (CR) areas, 2023

Code	Name	Randstad	Code	Name	Randstad
01	Oost-Groningen	No	21	Aggl. Haarlem	Yes
02	Delfzijl en omgeving	No	22	Zaanstreek	Yes
03	Overig Groningen	No	23	Groot-Amsterdam	Yes
04	Noord-Friesland	No	24	Het Gooi en Vechtstreek	Yes
05	Zuidwest-Friesland	No	25	Aggl. Leiden en Bollenstreek	Yes
06	Zuidoost-Friesland	No	26	Aggl. 's-Gravenhage	Yes
07	Noord-Drenthe	No	27	Delft en Westland	Yes
08	Zuidoost-Drenthe	No	28	Oost-Zuid-Holland	Yes
09	Zuidwest-Drenthe	No	29	Groot-Rijnmond	Yes
10	Noord-Overijssel	No	30	Zuidoost-Zuid-Holland	Yes
11	Zuidwest-Overijssel	No	31	Zeeuwsch-Vlaanderen	No
12	Twente	No	32	Overig Zeeland	No
13	Veluwe	No	33	West-Noord-Brabant	No
14	Achterhoek	No	34	Midden-Noord-Brabant	No
15	Arnhem/Nijmegen	No	35	Noordoost-Noord-Brabant	No
16	Zuidwest-Gelderland	No	36	Zuidoost-Noord-Brabant	No
17	Utrecht	Yes	37	Noord-Limburg	No
18	Kop van Noord-Holland	No	38	Midden-Limburg	No
19	Alkmaar en omgeving	No	39	Zuid-Limburg	No
20	IJmond	No	40	Flevoland	No

Randstad is yes if the COROP area is part of the Randstad, and no if the COROP area is no part of the Randstad.

Agglomeratie is abbreviated to Aggl.

Sources: CBS (2022), CBS (2023a).

Table B2 Overview of soil types included in the variable green space

Soil types in green space	Areas included
Agricultural surface	Areas used for greenhouses and other agricultural related usage such as grassland and areas for cultivation of cattle breeding and fruits.
Recreational surface	Public green space such as parks and greenbelts, spaces used for sports, community gardens, day recreational areas (such as amusement parks and zoos), and recreational accommodation areas (for overnight stays).
Woodland and nature surface	Woodland, dry natural open areas (such as beaches, dunes, and grassland not in use for agriculture), and wet natural open areas (such as wet heather).

Source: CBS (2023a).

Table B3 Overview of educational level classifications

Educational level	Description
Low	The highest obtained educational level is primary education, this includes all primary education, pre-vocational secondary education (VMBO), the first three years of senior general secondary education (HAVO) and pre-university education (VWO), secondary vocational education level 1 (MBO 1), and its predecessors.
Middle	The highest obtained educational level is secondary education, this includes the last 2 or 3 years of, respectively, senior general secondary education (HAVO) and pre-university education (VWO), secondary vocational education level 2, 3 and 4 (MBO 2, 3, and 4), and its predecessors.
High	The highest obtained educational level is higher education, this includes higher professional education (HBO), university education (WO), and its predecessors.

Source: CBS (2014).

Appendix C | Detailed summary statistics

Table C1 Descriptive statistics per variable per model

Variable per model	Obs.	Mean	Std. dev.	Minimum	Maximum
Green surface (in hectares)					
Total	360	71,681.56	42,880.96	7,735.00	164,334.00
Randstad	99	35,683.89	32,279.02	7,735.00	109,947.00
Non-Randstad	261	85,335.85	38,283.51	9,511.00	164,334.00
Agricultural	360	57,043.06	33,770.74	2,122.00	125,381.00
Randstad	99	27,882.78	26,144.95	2,122.00	84,453.00
Non-Randstad	261	68,103.86	29,503.87	4,050.00	125,381.00
Recreational	360	2,438.48	1,529.19	337.00	7,853.00
Randstad	99	2,677.12	2,188.72	508.00	7,853.00
Non-Randstad	261	2,347.96	1,180.37	337.00	5,434.00
Woodland and nature	360	12,200.01	13,353.67	359.00	78,206.00
Randstad	99	5,123.99	5,643.20	359.00	21,161.00
Non-Randstad	261	14,884.02	14,419.40	542.00	78,206.00
Gross value added (in millions of euros)					
Total	360	13,039.23	14,590.70	1,056.00	100,385.00
Randstad	99	22,795.46	22,861.80	2,311.00	100,385.00
Non-Randstad	261	9,338.58	6,852.46	1,056.00	34,267.00
Business services	360	4,347.42	6,809.34	122.00	53,632.00
Personal services	360	5,533.34	6,129.47	320.00	38,803.00
Manufacturing	360	1,695.64	1,439.53	305.00	8,547.00
Investments in fixed assets (in millions of euros)					
Total	360	3,111.00	3,676.53	225.00	39,052.00
Randstad	99	5,321.62	6,025.31	646.00	39,052.00
Non-Randstad	261	2,272.49	1,553.42	225.00	7,088.00
Business services	360	1,409.43	2,309.78	27.00	33,433.00
Personal services	360	1,080.30	1,230.10	68.00	7,724.00
Manufacturing	360	311.34	327.69	45.00	2,266.00
Labour productivity (gross value added per FTE in euros)					
Total	360	70,796.95	15,260.13	37,396.12	129,923.95
Randstad	99	77,275.63	14,525.78	46,533.33	117,961.22
Non-Randstad	261	68,339.52	14,831.30	37,396.12	129,923.95
Business services	360	87,659.94	17,867.95	38,125.00	158,674.56
Personal services	360	56,769.24	11,321.33	32,370.13	88,128.55
Manufacturing	360	87,110.45	25,376.57	33,942.86	165,460.98
Agriculture, forestry, and fishing	360	57,576.37	20,362.37	15,000.00	135,913.97

Labour volume (in thousands of FTEs)					
Total	360	171.80	156.45	14.70	851.00
Randstad	99	272.92	235.82	49.20	851.00
Non-Randstad	261	133.45	86.40	14.70	369.40
Business services	360	43.39	53.11	2.30	338.00
Personal services	360	90.93	84.13	7.80	440.30
Manufacturing	360	19.29	14.06	2.70	64.60
Educational level high (in percentage of employed people with high obtained education)					
Total	280	0.28	0.08	0.11	0.50
Randstad	77	0.34	0.08	0.18	0.50
Non-Randstad	203	0.26	0.06	0.11	0.45

Table C1 shows the summary statistics for all models. Total stands for the general model and the other models are labelled according to the industry names in Table 1. Total is divided into subsamples for the Randstad and non-Randstad models. The descriptive statistics of green soils are not given per industry as this variable does not vary across industries, only across COROP areas. The educational level high variable is not available for the years 2015 and 2017, and not given per industry, as this variable is only available on COROP level.

Appendix D | Multicollinearity checks

This appendix shows the different multicollinearity checks per analysis. All variables are put into one correlation matrix per analysis, however, either aggregate green space is included in the model or the three different types of green soil. These four green variables are never in one model together; therefore, the correlation between aggregate green space and (one of) the green soil type variables can be ignored. The Variance Inflation Factors are calculated for Pooled OLS regressions, and the dependent variable is the labour productivity of the model specification (i.e. general, industry specific, or region specific).

Table D1 Correlation matrix general variables

Variables	$Ln(G_{it})$	$Ln(A_{it})$	$Ln(R_{it})$	$Ln(W_{it})$	L_{it}	I_{it}	$E_{high,it}$
$Ln(G_{it})$	1.00						
$Ln(A_{it})$	0.97	1.00					
$Ln(R_{it})$	0.61	0.51	1.00				
$Ln(W_{it})$	0.75	0.60	0.68	1.00			
L_{it}	0.21	0.17	0.75	0.23	1.00		
I_{it}	0.22	0.18	0.74	0.22	0.97	1.00	
$E_{high,it}$	-0.27	-0.34	0.29	-0.01	0.51	0.51	1.00

Table D2 Variance Inflation Factors general models

Variables	Model 1	Model 2	Model 3
$Ln(G_{it})$	1.00	1.33	
$Ln(A_{it})$			2.35
$Ln(R_{it})$			6.42
$Ln(W_{it})$			3.00
L_{it}		17.85	19.42
I_{it}		18.18	18.71
$E_{high,it}$		1.72	1.96
Mean VIF	1.00	9.77	8.64

Table D3 Correlation matrix business services industry variables

Variables	$Ln(G_{it})$	$Ln(A_{it})$	$Ln(R_{it})$	$Ln(W_{it})$	L_{ibt}	I_{ibt}	$E_{high,it}$
$Ln(G_{it})$	1.00						
$Ln(A_{it})$	0.97	1.00					
$Ln(R_{it})$	0.61	0.51	1.00				
$Ln(W_{it})$	0.75	0.60	0.68	1.00			
L_{ibt}	0.11	0.08	0.64	0.11	1.00		
I_{ibt}	0.15	0.11	0.68	0.16	0.96	1.00	
$E_{high,it}$	-0.27	-0.34	0.29	-0.01	0.55	0.55	1.00

Table D4 Variance Inflation Factors business services industry models

Variables	Model 1	Model 2	Model 3
$Ln(G_{it})$	1.00	1.26	
$Ln(A_{it})$			2.30
$Ln(R_{it})$			5.26
$Ln(W_{it})$			2.97
L_{ibt}		14.44	14.76
I_{ibt}		14.91	15.97
$E_{high,it}$		1.77	2.03
Mean VIF	1.00	8.09	7.21

Table D5 Correlation matrix personal services industry variables

Variables	$Ln(G_{it})$	$Ln(A_{it})$	$Ln(R_{it})$	$Ln(W_{it})$	L_{ipt}	I_{ipt}	$E_{high,it}$
$Ln(G_{it})$	1.00						
$Ln(A_{it})$	0.97	1.00					
$Ln(R_{it})$	0.61	0.51	1.00				
$Ln(W_{it})$	0.75	0.60	0.68	1.00			
L_{ipt}	0.19	0.14	0.75	0.22	1.00		
I_{ipt}	0.19	0.16	0.69	0.17	0.95	1.00	
$E_{high,it}$	-0.27	-0.34	0.29	-0.01	0.53	0.48	1.00

Table D6 Variance Inflation Factors personal services industry models

Variables	Model 1	Model 2	Model 3
$Ln(G_{it})$	1.00	1.30	
$Ln(A_{it})$			2.36
$Ln(R_{it})$			6.78
$Ln(W_{it})$			3.01
L_{ipt}		11.96	15.82
I_{ipt}		10.87	11.44
$E_{high,it}$		1.75	1.94
Mean VIF	1.00	6.47	6.89

Table D7 Correlation matrix manufacturing industry variables

Variables	$Ln(G_{it})$	$Ln(A_{it})$	$Ln(R_{it})$	$Ln(W_{it})$	L_{imt}	I_{imt}	$E_{high,it}$
$Ln(G_{it})$	1.00						
$Ln(A_{it})$	0.97	1.00					
$Ln(R_{it})$	0.61	0.51	1.00				
$Ln(W_{it})$	0.75	0.60	0.68	1.00			
L_{imt}	0.45	0.40	0.72	0.47	1.00		
I_{imt}	0.32	0.27	0.60	0.35	0.86	1.00	
$E_{high,it}$	-0.27	-0.34	0.29	-0.01	0.14	0.15	1.00

Table D8 Variance Inflation Factors manufacturing industry models

Variables	Model 1	Model 2	Model 3
$Ln(G_{it})$	1.00	1.49	
$Ln(A_{it})$			2.31
$Ln(R_{it})$			3.94
$Ln(W_{it})$			2.22
L_{imt}		4.48	5.21
I_{imt}		3.87	3.89
$E_{high,it}$		1.19	1.71
Mean VIF	1.00	2.76	3.21

Table D9 Correlation matrix Randstad region variables

Variables	$\ln(G_{it})$	$\ln(A_{it})$	$\ln(R_{it})$	$\ln(W_{it})$	L_{it}	I_{it}	$E_{high,it}$
$\ln(G_{it})$	1.00						
$\ln(A_{it})$	0.95	1.00					
$\ln(R_{it})$	0.82	0.69	1.00				
$\ln(W_{it})$	0.54	0.29	0.70	1.00			
L_{it}	0.74	0.66	0.93	0.55	1.00		
I_{it}	0.73	0.65	0.91	0.52	0.97	1.00	
$E_{high,it}$	-0.02	-0.18	0.30	0.42	0.35	0.35	1.00

Table D10 Variance Inflation Factors Randstad region models

Variables	Model 1	Model 2	Model 3
$\ln(G_{it})$	1.00	2.82	
$\ln(A_{it})$			3.00
$\ln(R_{it})$			16.08
$\ln(W_{it})$			3.02
L_{it}		21.44	27.56
I_{it}		19.95	20.21
$E_{high,it}$		1.44	1.95
Mean VIF	1.00	11.41	11.97

Table D11 Correlation matrix non-Randstad region variables

Variables	$\ln(G_{it})$	$\ln(A_{it})$	$\ln(R_{it})$	$\ln(W_{it})$	L_{it}	I_{it}	$E_{high,it}$
$\ln(G_{it})$	1.00						
$\ln(A_{it})$	0.97	1.00					
$\ln(R_{it})$	0.67	0.54	1.00				
$\ln(W_{it})$	0.72	0.56	0.81	1.00			
L_{it}	0.46	0.34	0.79	0.63	1.00		
I_{it}	0.47	0.35	0.80	0.61	0.94	1.00	
$E_{high,it}$	0.07	0.00	0.38	0.27	0.50	0.56	1.00

Table D12 Variance Inflation Factors non-Randstad region models

Variables	Model 1	Model 2	Model 3
$Ln(G_{it})$	1.00	1.38	
$Ln(A_{it})$			1.63
$Ln(R_{it})$			5.36
$Ln(W_{it})$			3.12
L_{it}		9.49	10.07
I_{it}		10.82	11.33
$E_{high,it}$		1.58	1.58
Mean VIF	1.00	5.82	5.52